MASTERARBEIT\MASTER'S THESIS

Titel der Masterarbeit\Title of the Master's Thesis

„Using Recognition Data from Viennese Park-goers to Forecast UEFA Euro 2020 Football Matches"

verfasst von\submitted by

Ian Lundgren

angestrebter akademischer Grad\in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022\Vienna 2022

Studienkennzahl lt. Studienblatt\degree program code as it appears on the student record sheet:	UA 066 914
Studienrichtung lt. Studienblatt\degree program as it appears on the student record sheet:	Masterstudium Internationale Betriebswirtschaft
Betreut von\Supervisor:	Assoz. Prof. Steffen Keck, PhD

1 Abstract

The recognition heuristic holds that objects which are ”better” in a given regard are more well known. Sports players which are better are more well known, just as cities which are larger are more well known. One can thus use the fact that something is recognized as valuable information about its attributes; in sports this means that a more recognized team is more likely to win. I use the team recognition as well as the average recognition rate of all players on each team to forecast matches in the UEFA Euro 2020 football tournament. Three experiments are conducted where participants are grouped based on knowledge of football. In the first, respondents self-report their level of knowledge of football. In the second, participants self-report how often they view football and in the third participants are sorted by the number of players that they recognized. Forecasts will then be compared against the UEFA team rankings and aggregated betting odds. The results show a clear trend of more knowledge of football leading to better forecasts in all three trials for the player-level analysis while team-level analysis showed no clear trends. In the player-level analysis, the highest knowledge level performed as well as or nearly as well as the UEFA rankings, while the aggregated betting odds performed slightly better than all samples.

1.1 Abstrakt

Die Wiedererkennungsheuristik besagt, dass Objekte, die in einer bestimmten Hinsicht ”besser” sind, bekannter sind. Beispielweise sind bessere Sportler bekannter, ebenso wie größere Städte. Man kann die Tatsache nutzen, dass etwas über den zu messenden Kriteriumswert bekannt ist. Im Sport bedeutet dies, dass eine bekanntere Mannschaft mit größerer Wahrscheinlichkeit gewinnen wird. Ich verwende die Mannschaftsbekanntheit sowie den durchschnittlichen Bekanntheitsgrad aller Spieler jeder Mannschaft, um die Spiele des Fußballturniers UEFA Euro 2020 vorherzusagen. Es werden drei Experimente durchgeführt, bei denen die Teilnehmer*innen nach ihrem Fußballwissen gruppiert werden. Im ersten Experiment geben die Befragten selbst an, wie gut sie sich mit Fußball auskennen. Im zweiten geben die teilnehmenden selbst an, wie oft sie Fußball schauen, und im dritten Experiment werden die Teilnehmenden nach der Anzahl der von ihnen erkannten Spieler sortiert. Die Prognosen werden dann mit den UEFA-Team-Ranglisten und den aggregierten Wettquoten verglichen. Die Ergebnisse zeigen den klaren Trend, dass mehr Fußballwissen zu besseren Prognosen in allen Versuchen für die Analyse auf Spielerebene führt, während die Analyse auf Mannschaftsebene keine klaren Trends zeigt. Bei der Analyse auf Spielerebene schnitt das höchste Wissensniveau genauso gut oder fast genauso gut ab wie die UEFA-Rangliste, während die aggregierten Wettquoten etwas besser abschnitten als alle Stichproben.

1.2 Resumen

La heurística del reconocimiento sostiene que los objetos que son ”mejores” en un aspecto determinado son más conocidos. Los mejores jugadores son más conocidos, al igual que las ciudades más grandes. Así, se trata de reconocer algo sobre el valor del criterio que se está midiendo; en los deportes esto significa que un equipo más reconocido tiene más probabilidades de ganar. Utilizo el reconocimiento del equipo, así como el índice de reconocimiento medio de todos los jugadores de cada equipo para pronosticar los partidos del torneo de fútbol UEFA Euro 2020. Se realizan tres experimentos y se divide a los participantes según sus conocimientos de fútbol. En el primero, los encuestados autoinforman de su nivel de conocimiento del fútbol. En el segundo, los participantes autoinforman sobre la frecuencia con la que ven el fútbol y en el tercero se clasifican por el número de jugadores que han reconocido. A continuación se comparan los pronósticos con las clasificaciones de los equipos de la UEFA y las cuotas de apuestas agregadas. Los resultados muestran una clara tendencia a que un mayor conocimiento del fútbol conduzca a mejores pronósticos en las tres pruebas para el análisis a nivel de jugador, mientras que el análisis a nivel de equipo no mostró ninguna tendencia clara. En el análisis a nivel de jugador, el nivel de conocimiento más alto rindió tan bien o casi tan bien como la clasificación de la UEFA, mientras que las probabilidades de apuestas agregadas rindieron ligeramente mejor que todas las muestras.

2 Introduction

By far the most popular sport in the world, football is enjoyed by billions of spectators annually. This enormous popularity only seems to be growing and the most popular events in the world of football are the international competitions such as the World Cup or UEFA European Championships, colloquially referred to as the Euros. The governing body of football *Fédération Internationale de Football* claimed that over half of the world watched games during the 2018 World Cup in Russia (FIFA, n.d.-b). During these tournaments the people of the world come together to cheer for their team or against another.

Knowing that ”your team” is going to win is one of the most important issues for fans at football matches. While previously thought that uncertainty of outcome was a significant driver of attendance at sports games, research has shown that uncertainty is not as important as previously thought (Pawlowski & Anders, 2012). In their work, (Buraimo & Simmons, 2008) found evidence even more contrary to popular belief in economics; they found evidence that supporters would rather see their teams beat weaker teams than play highly contested matches based off of data from English Premier League.

What fans really want to know is will their team win the match. During large matches, millions of liters of beer are consumed and absurdly expensive replica jerseys are dawned in the hope of helping their team to win.

Others go further, ritualizing their sports consumption activities both as an act of obtaining a cultural identity (Chun et al., 2005) as well as the more well known practices of routines in the hopes of bringing good fortune to their team.

In order to help determine if their team will win the match, fans have gone to great lengths to be able to ”predict” match winners. One of the most famous examples of this is Paul the Octopus. He was used to predict the matches for the German National Team during the 2008 Euros and the 2010 World Cup, gaining cult-status (Christenson, 2010). During the 2018 World Cup this trend was continued with an elephant, a polar bear, and a pig being used to ”predict” match outcomes (TRTWorld, 2018).

There are generally more widely accepted ways of predicting match outcomes. One of the most often used is that of the expert opinion. Often on TV broadcasts before and after each match there will be former coaches or players who are there to offer their opinions on what they think will happen before the game and what they thought about what happened after the game. The thought is that they will leverage the past experience and expertise in the sport to be able to better predict what the likely outcome is. Despite their years, or often decades, of experience experts often are incorrect with their predictions.

Another often-used method for predicting sports outcomes is by the ranking that each player or team has. There are many different ways that these rankings are calculated and each sport uses different models. Most of these take into account the past performance in order to try to predict the future outcomes. The models used range from averaging expert rankings in US college (American) football to relatively simple ranking such as the Elo system used in chess which takes into account the outcome and the match and the ranking of the other player. FIFA has recently announced that they will change their ranking system to one similar to this where the importance of the game and the difference in the team rankings will be factors in determining team ranking (FIFA, n.d.-a).

By far the most complex method of predicting team outcomes is done in the sports betting world. Each of the bookmakers have small armies of statisticians who's job it is to use vast amounts of data in order to get a better insight into what the outcome is likely to be. The betting market facilitates discovery in the same way that the stock market in theory allows all participants to collect and analyze data to make the most informed decision about the true price of a good.

In recent decades a the field or research surrounding the recognition heuristic has shown itself to be able to provide surprisingly accurate forecasts, and with only a fraction of the information needed for bookmakers predictive models. The premise of the work at its core is that knowledge is not random, but is structured by the relative frequency that one encounters any item. It is likely that one has heard of Stockholm but not of Örebro, and by this fact alone it would be safe for someone asked which of the two cities is larger to say Stockholm by the mere fact that they recognize it and not Örebro.

Research surrounding the recognition heuristic started off in the field of decision making. In tasks nearly identical to this they asked participants to choose which of two cities is larger. In their work Gigerenzer et al. (1991) took the 65 cities in Germany with over 100.000 inhabitants and asked participants to guess which of the two cities was larger for each of the possible 300 pairings. In this, participants correctly chose the larger city 74% of the time.

In their paper, (Scheibehenne & Bröder, 2007) were the first to apply the recognition heuristic to sports prediction in their study predicting outcomes in the Wimbledon 2005 Tournament. They asked both laypeople and amateur tennis players to predict match outcomes and found that the participants were able to predict matches as well as or better than the ATP tennis ranking system. More recently, work has used the ”atomized recognition rate” to predict matches in the World Cup 2006 and UEFA Euros 2008 (Herzog & Hertwig, 2011). In this work I aim to test the predictive capabilities of Viennese park-goers using the model developed by Herzog and Hertwig as well as looking for evidence of a less is more phenomenon.

3 Literature Review

3.1 Origins of the Recognition Heuristic

The beginnings of research into the recognition heuristic came from work into confidence of decision making. In their work in this field, Gigerenzer and Goldstein (1991) sought to study the overconfidence effect (the fact that people routinely estimate that their guesses are correct at a higher rate than the actual correct rate of guesses) and the hard-easy effect (which finds that people overestimate the likelihood of being correct for difficult tasks while underestimating the likelihood that they are correct for easy tasks). Gigerenzer and Goldstein listed the ”familiarity cue” (if one had heard of the city before) as one of the factors that might influence confidence levels.

This insight helped to explain the rather counter intuitive results that Ulrich Hoffrage got in his doctoral dissertation. In his paper (Hoffrage, 2011), he discusses the research from his unpublished dissertation. In it, both German and US students were asked to participate in a study similar to that of Gigerenzer and Goldstein (1991) in which they were asked to indicate which of two German city pairs were larger. Astonishingly, the US students performed slightly better than the German students at guessing which of two German cities was larger.

As recalled in (Gigerenzer & Goldstein, 2011), this completely ruined the study, as they could not figure out how it was that US students who knew less about German cities could perform as well (or even better!) than the German students who knew more about German geography. They attribute the fact that this went so long without being discovered in research to the fact that regression modelling is additive. When adding an additional explanatory variable, this never decreases the the fit of a model; this ”more is better” principal has been one of the basic tenants of regression model building.

In their paper, Goldstein & Gigerenzer (2002) point out that it is possible for a situation to arise in which less knowledge could be adventitious. They point out that the recognition heuristic is helpful for decision making when the criterion is correlated with recognition. The work by Golstein & Gigerenzer also propose recognition validity, α as being calculated by:

$$\alpha = R/(R + W), \tag{1}$$

In this equation, R and W are the number of right, ”correct”, and wrong, ”incorrect” respectively. This is calculated for all pairs in which one item is recognized but not the other. Knowledge validity β is the proportion of correct answers where both of the options are recognized.

This can lead to a situation in which less knowledge could be more advantageous, which has been deemed the ”less is more effect”. It has been shown that the less is more effect will arise in situations when $\alpha > \beta$ (Goldstein & Gigerenzer, 2002). The effect gets its name from the fact that the proportion of answers that are correct using the recognition heuristic is higher than the proportion of correct answers given using knowledge. There has been some research conducted by Smithson (2010) which has demonstrated that the less is more effect can occur in cases where $\alpha \geq \beta$ and may also not arise in cases where $\alpha > \beta$ though.

One of the major debates in the research around the recognition heuristic is whether it is a non-compensatory decision making strategy. Non-compensatory is given to mean that other information is not considered when using the recognition heuristic (Goldstein & Gigerenzer, 2002). Goldstein & Gigrenzer demonstrated this by asking a about German city pairs and then giving additional information which they instructed the participants were indicative of city size. Despite being told that these ques had strong correlations with city size, they found evidence that suggests participants did not take this additional information into consideration.

Several studies have found evidence which contradicts the claims of recognition as a non-compensatory decision making strategy. In a study at Stanford, participants were asked which of city pairs were larger. The city pairs were of a small, well known city and a made-up city in the first part of the study. In the second part, the fictitious cities were replaced with places known for attributes other than size (ie. nuclear disaster) (Oppenheimer, 2003). They found that participants did partially rely on recognition for choosing the city but it was also clear that recognition was not the only process at work. If this were the case then participants should have chosen the nearby but small or famous places for non-size attributes, over the fictitious cities that by definition could not have been recognized.

Newell and Shanks (2004) used a mock stock market in order to investigate if additional information is incorporated into the decision making process. They created a mock investment situation with fictitious companies and financial ”advisors” who’s advice could be purchased. Participants were told that the ”advisors” were not equally good and were primed that recognition was a valid cue for stock performance. Four stocks were presented repeatedly to the participants so that familiarity could be gained. Despite this, they found that participants used the advice from the ”most valid advisor” at a higher rate when recognition was low.

In their article, Bröder and Eichler 2006 used a city choice model similar to those previously used. Participants were presented with paired fictitious city names, some of which had been primed for recognition by appearing five times in the experiment. During the experiment, cues were given that indicated larger city size; these were if a city had an exposition site, intercity train line, or major league soccer team. Each additional positive cue for ”recognized” cities had a significant impact in increasing the likelihood that it was chosen by participants; lending further evidence in support of the hypothesis that recognition is not the only factor at play when making decisions. One issue surrounding such work is that it has not been shown that recognition gained within an experiment has the same validity as recognition gained ”naturally”.

Further research by Pohl 2006 also provided supporting evidence that recognition cues are important but that they are influenced by further information. Their research also demonstrates an important aspect of heuristics used in the decision making process. In experiment one of (Pohl, 2006) half of the participants were asked to compare city pairs based on their population size while the other half were asked to which of two cities was closer to Interlaken, a city close to the geographical center of Switzerland. 89% of participants in the city size group selected in accordance with the recognition heuristic while in the distance group only 54% of guesses aligned with recognition.

The difference between the rates at which respondents' answers agreed with recognition demonstrates an inherent understanding of ecological rationality. This was given by (Goldstein & Gigerenzer, 2002) to mean that one is able to exploit structures in the information in the environment. Much in the same way that a professional billiard player is not likely to be able to pass a physics exam yet has a very good understanding of Newtonian physics, people, although unaware even what the recognition heuristic is, seem to gave a good idea of when the recognition heuristic is applicable.

3.2 Applied Uses of the Recognition Heuristic

Herbert Simon described bounded rationality as a pair of scissors, the two blades of which are the cognitive limitations of humans and the structure of the environment. It has been suggested that we make use of an adaptive toolbox of simplified strategies and heuristics to allow us to more quickly make decisions (Gigerenzer & Todd, 1999). A study of Dutch database marketing companies found that a simple rule of thumb can work better than more complex models (Wübben & Wangenheim, 2008). A simple heuristic was used at a marketing agency. If a customer had not made a purchase for more than 9 months, then they were to stop receiving mailers. They found that this simple rule predicted whether customers were likely to make a return purchase better than more complex models.

The recognition heuristic has been shown to be a good predictor of election results. In Gaismaier & Marewski (2011) they studied three elections in Germany, two state and one federal for party recognition and voter intention. The samples used were ”convenience” samples of passers-by; which they note would be considered bad samples by pollsters. On top of this the three experiments had sample sizes of fifty nine, sixty six, and thirty four participants. Despite these limitations, their study proved surprisingly effective at predicting election outcomes; this was particularly true for small parties.

One thing that fast and frugal heuristics do exceptionally well is extrapolating to larger data sets. More complex models often fall into the trap of being fit to the data set that they are mirroring rather than the population values; as the complexity of a model increases so does the likelihood that it is over fitting to the data set.

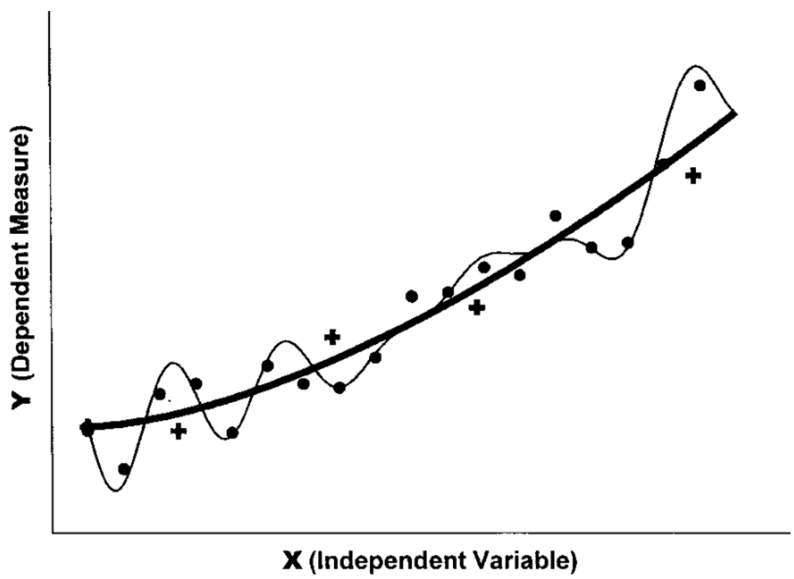

Figure 1: This figure taken from Pitt et al. (2002) demonstrates the trade off in model complexity. The more complex a model is, the better the goodness of fit, all else equal. The thin line represents a model which has a very high goodness of fit given the data set of circular points. With the addition of the plus-shaped data points, the simpler model represented by the thick line is more appropriate. When there is a gap between goodness of fit and generalizability the model will suffer from over fitting. See (Pitt et al., 2002) for further discussion.

Because fast and frugal methods are better at producing models which have higher generalizability and are not as prone to over fitting they are being applied in an increasing number of fields. This includes stock picking (Andersson & Rakow, 2007), medicine (Wegwarth et al., 2009), and the law (Gigerenzer & Engel, 2006). For a more complete discussion of applied heuristics see (Hafenbrädl et al., 2016) and (Marewski et al., 2010).

3.3 Recognition Heuristic in Sports

The recognition heuristic is ecologically valid in the sports domain because sports is discussed in the media as well as the public domain. Even those who do not watch tennis likely know who Roger Federer or Serena Williams are just as those who don't watch football are likely to know to who Lionel Messi or Christiano Ronaldo are. Players that are good are discussed in the media as well as in social interactions and thus become known by the population generally. The better a player is, the more likely they are to be recognized by any given person. Thus, players who are better will have a higher recognition validity.

Some of the first research into the recognition heuristic in sports backed up these conclusions. Snook and Cullen (2006) asked participants to choose between paired hockey players, to see if the participants would be able to choose the player with higher career points. They found that the recognition heuristic was adhered to 95% of the time and were able to choose the player with higher lifetime points in 94% of cases. Higher rates of players recognized correlated with better performance, although interestingly, women recognized fewer players, but were not less accurate in their predictions.

Serwe and Frings (2006) used the recognition heuristic to predict matches at the 2005 Wimbledon Men's tennis tournament. They collected recognition data for players in the tournament from a group of 29 amateur tennis players as well as 96 students (laypeople). The participants were asked to indicated if they had ”heard of” a player before or not. The players in the tournament were then ranked by number of times that there were recognized by each group of participants. This ranking system performed exceptionally well, with armature players' recognition-based ranking outperformed the ATP (the Association of Tennis Professionals) entry and championship rankings while the laypeople did as well as the ATP entry rankings but were outperformed by the ATP championship rankings. In a similar study from the 2005 Wimbledon Men's Tennis Tournament Scheibehenne and Bröder (2007) collected player name recognition data form amateur tennis players and laypeople. Again, the predictions based on the recognition data performed as well or better than the ATP rankings and the tournament entry seeds.

A study of Turkish and English students' ability to predict English F.A. (Football Association) 3rd round matches (Ayton et al., 2011) found a similar ability to predict matches between English and Turkish students. Participants were asked to choose which team they though would win each draw (match, of which there are 32 in the 3rd round) and the percent confidence in this prediction 50-100%. The FA cup is open to professional as well as non-league clubs which consist of semi-professional and armature clubs; as such, Turkish participants recognized fewer of the participating clubs. Despite recognizing fewer clubs, Turkish students had prediction rate similar to that of English students, providing evidence in support of a less is more effect.

A study of Swedish students and football experts (Andersson et al., 2005) found that simply asking for match predictions is less accurate at predicting than using the recognition heuristic. For the 2002 World Cup they asked students (166 Swedish and 41 US-American) as well as 52 (all male) experts to predict which two teams they thought would advance out of the group stage. There was little difference in the prediction accuracy between the groups; both experts and laypeople performed little better than chance at predicting teams to advance. Similarly, for the 2006 World Cup it was found that experts and laypeople performed equally well in predicting advancement to the second round (Andersson et al., 2009).Just as in 2002, participants performed better than chance in predicting advancement in the tournament while most participants were outperformed by advancement based on team rankings.

During the 2004 European Football Tournament Pachur and Biele (2007) conduced a recognition-based study to test prediction abilities of 121 laypeople and 20 experts. Each participant was asked to forecast which two teams they though would go through to the knockout stage of the tournament (forecasting) as well as asked which of the teams that the recognized (recognition). They did not find evidence for a less is more effect, but the opposite; laypeople predicted 64.7% of matches correctly while experts did better, predicting 76.6% correctly.

Working off of the framework of player-based recognition from the tennis model in (Serwe & Frings, 2006), Herzog & Hertwig (2011) applied this to the 2006 World Cup and the 2008 UEFA Euro. By ranking the teams in the competition by the average player recognition rate for that team they were able to leverage the recognition heuristic as well as the wisdom of the crowd (Galton, 1907). This player-based recognition ranking Herzog & Hertwig call ”atom recognition rate” since the players on the team are the smallest whole unit of analysis. This method also has the added benefit of being able to predict matches past the first round. In order to predict past the first round using a head-to-head comparison one would need to gather data for every possible combination of teams that could play each other. For a tournament such as the World Cup with 32 participating teams, it would be unreasonable to assume that any participant would fill out the entirety of a questionnaire that takes an hour.

For the World Cup and UEFA Euro 113 and 517 respondents respectively judged a random third of the players in the tournament. For the World Cup study, all participants were Swiss while for the Euro a plurality were

Swiss with the next larges group being Germans. In both cases, the player-based recognition predictions performed exceptionally well. For the World Cup 2006 participants predicted 84% of games correctly while for the Euro 2008 they predicted 62% of games correctly. In both cases, the recognition heuristic outperformed predictions based off of FIFA rankings while being outperformed by the betting odds.

4 Hypothesis and research questions

In this paper, I will use the atom recognition rate developed by Herzog & Hertwig as well as team recognition rate to forecast the 2020 (2021) UEFA Euro Football Tournament and test how the knowledge level effects the forecasting ability of participants. By increasing the sample size I aim to be able to capture enough responses at all levels to be able to compare performance across knowledge levels within one sample, instead of selecting different samples to represent each knowledge level. Additionally, this will allow a more detailed analysis of performance by knowledge level than has been conducted using the atom recognition rate. I will then be able to check for evidence of a less is more effect, or if this is not found, determine the knowledge level at which participants most accurately predicted match outcomes.

In line with the findings from Serwe and Frings (2006) and Schiebenhenne and Bröder (2007) who both found that increased knowledge lead to better predictions in tennis tournaments. Since they collected individual player recognition information instead of simply team recognition data, this is most similar to the atom-recognition model where recognition data is collected for each football player. If there is a relationship between between football knowledge and forecasting performance I will observe at what level it best performs in order to better direct future research to focus on this knowledge level.

5 Method

5.1 Participants

During the four days prior to the kickoff of the UEFA Euro 2020 I approached people in parks to ask them to participate in a survey about football knowledge. After cleaning the data to remove responses in which no players were recognized and one in which every player was recognized I was left with 788 complete responses. The gender divide of respondents was fairly equal, 51.8% were female, while 46.1% were male and the remaining 2.2% preferred not to say. The age of of participants was slightly younger with an average age of 25.3 with a standard deviation of 6.6; the youngest participant was 13 and the oldest was 67. The fact that a majority of respondents are students reflected their slightly younger section of the population. Those who reported their primary status as students represented

57.2% while those working represented 37.2% and all ”other” were 5.6%. Given that travel was still relatively restricted due to the COVID-19 pandemic at the time of the survey I expected that the majority of respondents would be from Germany and Austria. This bore itself out in the data with 60.8% of respondents reporting themselves as being of Austrian citizenship while 17.9% identified as having German citizenship. The remaining portion comprised the rest of the world with the next largest portion after Germans being Italians at a much smaller 4.6%.

Included are two charts showing the self-reported knowledge levels of the participants. In both cases the options were presented as a five-point likert scale. The first of which depicts how frequently participants report watching football. As anticipated, there is a skew towards increasing knowledge in the data. A majority of participants reported watching football ”never” or ”during major tournaments” while those who reported viewing football ”once a month”, ”weekly”, or ”more than once a week” represented just under one quarter of those surveyed. For the self-reported knowledge level of football the knowledge skew was even more pronounced. Those reporting a ”poor” knowledge of football represented a plurality of respondents. Each increasing level of knowledge is the next largest with ”excellent” representing the smallest group.

Because Austrians and Germans represent such a large portion of the sample this could be problematic for the results. This would be because Germans and Austrians would know more players on their national teams due to cultural proximity rather than the players being better. Just as Finnish media more likely to discuss Finnish football more than another country's, German and Austrian media will also discuss their own (and each other's, due to the shared cultural tenants) football teams and players. This would result in a higher recognition score than we would otherwise expect for players of the German and Austrian teams (since both are in the tournament) than we would otherwise expect. If this is true, then we will see an overestimation of the performance of Germany and Austria and thus worse results compared with responses from non-German speakers.

In order to counter this, I will remove Germans and Austrians from answering about their own teams. Additionally since the cultural landscapes are so similar and there is a shared language, this is also likely to be true for the players on the other nations' team. As a point, the largest the German and Austrian cultural landscapes are so similar. Additionally, Germans make up the largest group of foreigners in Austria (von Martin Mohr & 17.02.2022, 2022) and thus it is quite likely that if someone identified most as German on the questionnaire, they may be living in Austria and be culturally identical to an Austrian.

5.2 Questionnaire

During the four days prior to the start of the Euro 2020, park goers in several Viennese parks were asked to participate in the survey. First, they were

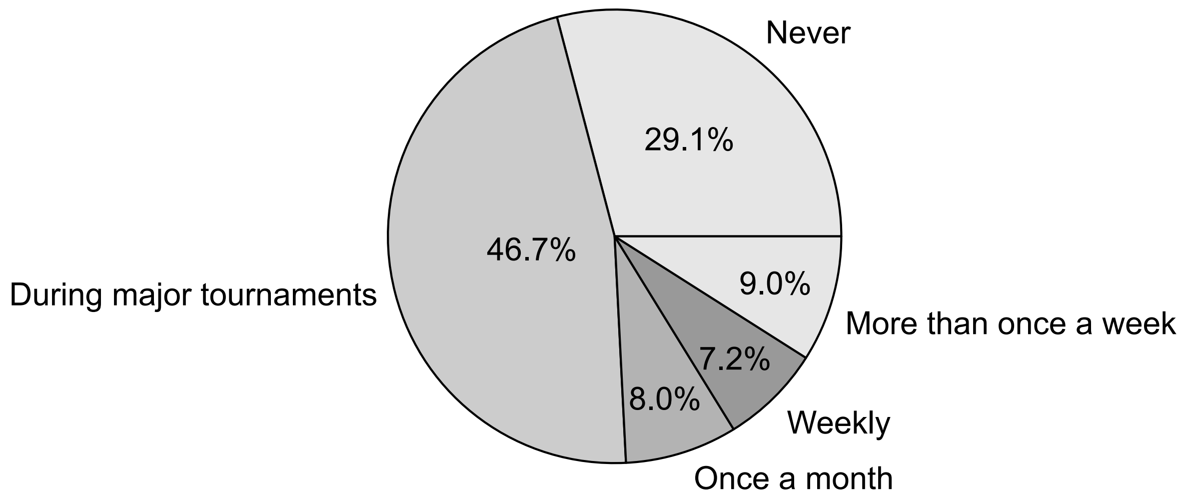

Figure 2: The frequency with with participants reported watching football

Figure 3: The self-reported knowledge level of participants

approached and explained that the survey was for my master's thesis. For those that showed interest, it was further explained that I was studying the wisdom of the crowd ”Schwarmintelligenz” in the context of the Euro 2020. The most common question was if this was a trick questionnaire to see if participants ”recognized” football players who do not exist. They were informed that this was not the case, but that the questionnaire sought only to see who the most recognized players in the tournament were.

The survey was administered online and accessible via a QR code presented after they had agreed to partake in the survey. Once scanned the QR code linked to a github project which randomly decided between the tree versions of the questionnaire. The code generated a random number between 0 and 1, then multiplied this by three. The resulting value dictated the link that participants would be shown upon scanning the QR code. The survey was broken up into three groups with each group containing a random third of players. This was chosen because of the high number (622) of players in the tournament. It was thought that if respondents had to give recognition data fora all players in the tournament this would lead to a very low completion rate.

The time to complete the survey was not collected as many participants did not finish it immediately. Once the QR code was scanned one of the three versions of the survey would open in a tab on the phone's browser. This would stay open in the browser tab and could be completed later; a choice taken up by many who did not wish to interrupt their time in the park. From anecdotal evidence of participants who I later ran into after having completed the survey, it was estimated to take about 10 minutes to complete.

Canvassing generally took place from 10am to 10pm each of the four days with the exception of the day of the start of the tournament. The ability to submit responses was cut off at one hour before kickoff of the first match. After this time pregame show would have started and was likely to introduce participants to players who the otherwise would not have known. The main canvassing areas for the survey were in Augarten, Donaukanal, and Donauinsel while some additional canvassing was done in the courtyards of the Altes AKH, as well as Volksgarte, Heldenplatz, and Burggarten although the responses gathered at the latter grouping represent a significantly smaller sum than those from the former. During the morning and midday, surveys were generally collected from the Donauinsel where the participants tended to be students studying for exams in the sun. During the afternoon and evening surveys were generally collected at Augarten and the Donaukanal. As these locations are closer to the city center they tended to have a higher proportion of working people coming out for a drink or walk and to enjoy the weather after work.

The survey was structured into three parts. The first of which was the collection of demographic data. Participants indicated their age, sex, nationality, knowledge of football, and how often they watched football. The second section consisted of team-level recognition data collection. Participants were presented with the team crests for each of the teams participating in the tournament and asked if they recognized each crest with the options being yes, no, and not sure. In the final section participants were asked about individual football players in the tournament in order to gather ”atom recognition” data. For this there were nine questions with approximately 23 each; participants could select any number of responses, including none. A ”I don’t recognize any of these players” option was provided at the bottom if this was the case.

5.3 Procedure

Ten days before the first match, the team rosters were finalized and published. Each team could select 26 to participate in the event. Every team took 26 players except for Spain which only selected 24 for reasons not given. Team participant information was gathered from the UEFA website (UEFA.com, 2021). The names of all participating players were organized in a spreadsheet with additional relevant information added. For each player a random number was generated and the list of players was then sorted from lowest to highest. In this order the list was split into thirds with each third representing one of the three survey groups.

After the responses were submitted, a spreadsheet was downloaded for each group. These were then combined in order to be easier to work with. The results were then analyzed three different ways in order to test performance by knowledge level. Firstly, a general analysis was made based off of all valid responses. This was followed by an analysis of the individual levels of knowledge based on the self reported knowledge and frequency of viewing football. Additionally, an analysis was conducted by sorting by the number of players recognized. Five equal sized groups were formed by ordering responses from least to most player recognized. I chose five groups since this is the same number of groups as in both other knowledge-based analysis. This was done as a measure to check if the self-reported knowledge levels coincided with how knowledgeable they really were about football (measured in the number of players recognized).

The number of instances that a players name was checked off as ”recognized” was summed for all participants in the group. This was then divided by the number of participants in that group. The scores of all players on each team were averaged to give the atom recognition rate for the team. These rates were then compared against each other in order to create a ranking. Higher atom recognition rates should (per the hypothesis) mean that a team will do better because it has more recognized (and thus good) players.

In order to determine how well the team as an entity is known I used the team recognition rate. For this the team crests were used to gauge the level of knowledge. The mechanism for recognition are the same for that based off of the players. The better a team is and thus the more that they are discussed, the more well known that team should be.

5.4 Validity of team rankings

In order to determine the effectiveness of the recognition-based prediction methods I will compare them to both the predictions of official team rankings and predictions of the betting odds market. Team rankings are widely used to predict outcomes in sports and their determinants differ across sports. As mentioned, US college football relies partially on an opinion-based system in combination with the teams ranking, but most other sports use more scientific methods.

One main benefit of the US college football ranking system relying on expert opinions rather than purely on team performance metrics is that they are able to create team rankings before matches have been played in a season. This is especially relevant in US college football because the composition of the team changes significantly each year as older players graduate and newer players replace them. This means that the strength of teams can vary greatly from season to season. The downside to this methodology is that it creates opacity in how the rankings are arrived at. Lebovic and Sigelman (2001) analyzed the ranking system and found several consistent trends in their analysis. Each win will lead to only a minor increase in rank while each loss leads to a significant decrease in rank and wins earlier in the season seem to count for less than wins later in the season.

Boulier and Stekler (1999) analyzed how well rankings (seeds) performed in predicting the outcome of professional tennis matches and US college basketball games. For this, they analyzed how the difference in rankings between competitors effects the expected outcome. They found evidence that as the number of ranks between player or teams increases so too does the chance of winning for the higher seeded team; for both sports the samples were several hundred games/matches. Notably for tennis there was no observed advantage for players who had one rank advantage over their opponent, although this was not the case for basketball. This could be due to the fact that tennis has a higher number of ranked players and thus the difference in skill between ranked units in any given system would decrease as the number or ranked units increased.

Likewise, Clarke and Dyte (2000) created a model to test in several tennis tournaments how well seed ratings would predict the chances of tennis players winning each round given their opponent. They were able to update the model each round to account for the players who had been knocked out and thus adjust the likelihood of winning. Generally, their model improved as the tournament went on, although during the 1998 US open they predicted the eventual winner with the highest probability of winning from start to finish.

More recently Corral and Prieto-Rodríguez (2010) created a probit model to test how different factors such as differences in rankings and physical characteristics effected tennis outcomes. They observed Grand Slam tennis matches form 2005 to 2008 for both men and women. For both men and women the difference in ranking was found to be an important determinant variable in the model. Their findings provided further evidence that differences in rank play an important role in determining tennis match outcomes.

The ranking system used for national teams in football is the FIFA/Coca-Cola World Ranking. This takes into account the past as well present performance of teams in order to measure their strength. Suzuki and Kazunobu (2008) analyzed how well FIFA rankings performed in the Worlds Cup 1994, 1998, 2002, and 2006. Teams which were ranked among the top 16 in the world were significantly more likely to advance to the final tournament (the part that is on tv – whereas the playing games to determine which teams get to participate are technically part of the tournament as well but not nearly as well known of). Those teams ranked among the top 16 over the time that the study was conducted had an average 72.9% chance of making it to the final tournament compared to an average of 31.3% for all teams not in the top 16. It was also noted by the authors that the rating system was chanced after the World Cup 2006. Until then games in the last eight years had been factored into the team ranking. This was changed since games played seven years ago likely have little bearing on the current state of the team. This chance thus caused the ranking to better reflect how the teams are currently performing.

An analysis of different methods of predicting the matches for the Euro 2008 by Leitner et al. (2010) also observed the predictive abilities of team rankings in international football. They found that the FIFA rankings did well in predicting outcomes. Furthermore they also found that the FIFA rankings were an attempted measure of the teams current potential. This would lend evidence to the fact that the rating reform following the 2006 World Cup helped to address the strong weight given to long past games.

5.5 Validity of betting odds

The other method of forecasting used by Leitner et al. (2010) was that of the betting odds in the online betting market. They found that the betting odds outperformed all other methods of forecasting that they observed. Thaler and Ziemba (1988) point out that betting markets are even better at testing market hypothesis than even the stock market. This is because, they argue, bets have a set time at which the value of it is known based on the outcome. This differs from the stock market in that there is no definite endpoint for the values of the assets to be calculated at. This should allow even greater integration of relevant information into the model and decision making process.

Where we expect financial markets to be efficient, see (Malkiel, 2003), we can thus expect betting markets to have an even lower level of arbitrage opportunities. Indeed, several papers looking into this have been able to find some very small market inefficiencies, spread over years of odds. Vlastakis et al. (2009) looked for arbitrage opportunities in betting on European football. They did this by looking for discrepancies in the odds between bookmakers which would allow them to place multiple bets across bookmakers which should result in an expected payout greater than 100% of the initial bets. In their study they observed 12,420 matches, in which only 63 had arbitrage opportunities. For those matches which were offered by online betting agencies (a requirement if one is to place multiple bets across different betting platforms) there were 10,374 matches of which only 10 contained arbitrage opportunities. This 0.0096% rate of matches presenting arbitrage opportunities demonstrates the degree of efficiency that exists in the betting odds market.

Another test of betting odds market efficiency was carried out by Gil and Levitt (2007) where they observed the betting market for the 2002 World Cup. They found an arbitrage strategy of buying contracts just before a goal was scored, then selling them again approximately 15 min after the goal had been scored. This, however, is only possible to do in hindsight, since one will not know when a goal will be scored. They also found that pre-game favorites had a negative return and that market-makers loose money but stay in the marketplace.

Further evidence for the predictive power of betting odds in sports has been put forward by Boulier and Steckler (Boulier & Stekler, 2003) who found that betting odds were the best predictor or game outcomes compared with several different methods analyzed, mostly based off of expert opinions. Forrest et al. (2005) found that for five years of English football games, betting odds performed better than bootstrap methods or statistical models. Recently some research has centered on the peer to peer betting markets (Franck et al., 2010) and using wisdom of the crowds among bettors to improve on betting outcomes (Brown & Reade, 2019). Despite moderate successes in both cases, neither were able to find any evidence of significant faults in the market efficiency of betting odds market.

Because of its high efficiency, betting odds are often used a point of comparison in the literature on sports forecasting. In the papers reviewed in the literature review which compared different findings to the betting odds, none outperformed the betting odds. So far, it has shown itself to be the gold-standard of sports forecasting. For this reason, I will be comparing the predictions made in this paper against those of the betting market. In order to do so, I use ("Odds portal - betting odds monitoring service", n.d.) which aggregates and provides an average of the odds for the 62 largest online betting exchanges around the world.

5.6 Matches

I will be observing the matches of the UEFA Euro 2020 men's football competition which took place during the summer of 2021. There are 24 teams competing in the tournament, each national team is placed in groups with three other national teams; this results in six groups of four teams each. Each team will play every other team in their group one time with three points being awarded for a win, one point for a draw, and none for a loss. The two teams from each group with the most points then will move on to the knockout stage of the tournament. In this stage, teams must eliminate their opponent in order to move on to the next round. A draw is not possible in this stage and if the game is tied after the 90 minutes of play then there will be two additional periods of extra play time. Each half of the extra time is 15 minutes and if there was still not a winner after the full 120 minutes then the match would go to a penalty shoot-out.

This means that in total there are 51 matches in the tournament. I will use matches for which there is a winner during the run of play; either during the 90 minutes of standard play or during the added 30 minutes which are possible starting in the knockout stage. Those matches which end in draws are not used because there is no winner so it can not be used in this analysis. Additionally, matches which have an eventual winner which is decided by a penalty shoot-out will not be used because the outcome of penalty shootouts are shown to have little relation to team skill (Jordet et al., 2007; Wunderlich et al., 2020).

6 Results

6.1 Rates of recognition

Below are three figures which show the distribution of player recognition for respondents. The first two figures display the number of players recognized within each self-reported knowledge level. In the first instance this is by the self-reported football knowledge level while in the second figure the self-report frequency of viewing football is displayed. Both figures show a clear positive correlation between self-reported knowledge and the number of players recognized. This is the expected result, as one would expect that as respondents identify as having more knowledge about football and view more matches they would be able to recognize a greater number of players. In particular the top two options for most knowledge and most frequent viewing show a significant increase in the number of players recognized compared to the lower rankings in each respect. Since it is shown that increased knowledge is correlated with increased player recognition this demonstrates that higher levels we can then test how each of these levels fairs in the prediction of matches in order to determine if lower levels can predict equally as well as the higher levels.

The third figure is the control sample where I sorted the responses into equal fifths for each group ranging from least to most recognized players. The first group contains responses in which the fewest players were recognized while the fifth group contains responses in which the most players are recognized. The variance in number of recognized players for each knowledge level is much greater in the two previous examples. This will allow the prediction ability to be tested on the basis of the number or recognized players. If it is indeed the number of players recognized which has a larger impact on the predictive ability of the model, then better results would be expected for the model using the data sorted by number of responses.

6.2 Recognition performance

As was expected, the recognition heuristic was able to predict a high percentage of the games of the UEFA Euro 2020 correctly. When observing the predictive abilities of the team recognition, this performed worse than the player based recognition forecasts, as expected. Interestingly, the portion of games that were predict corrected using this method do not seem to be correlated with the participants' knowledge of football. In the self-reported knowledge sample, there is a slight positive correlation with the data, while the frequency of viewership has a negative correlation and the equal fifths has no correlation. The same holds true when we observe the team-based recognition with German and Austrian participants omitted from answering about their or each others teams. For division of equal fifths sorted by number of players recognized, there is no discernible trend, while for both self-reported knowledge and frequency of viewership there

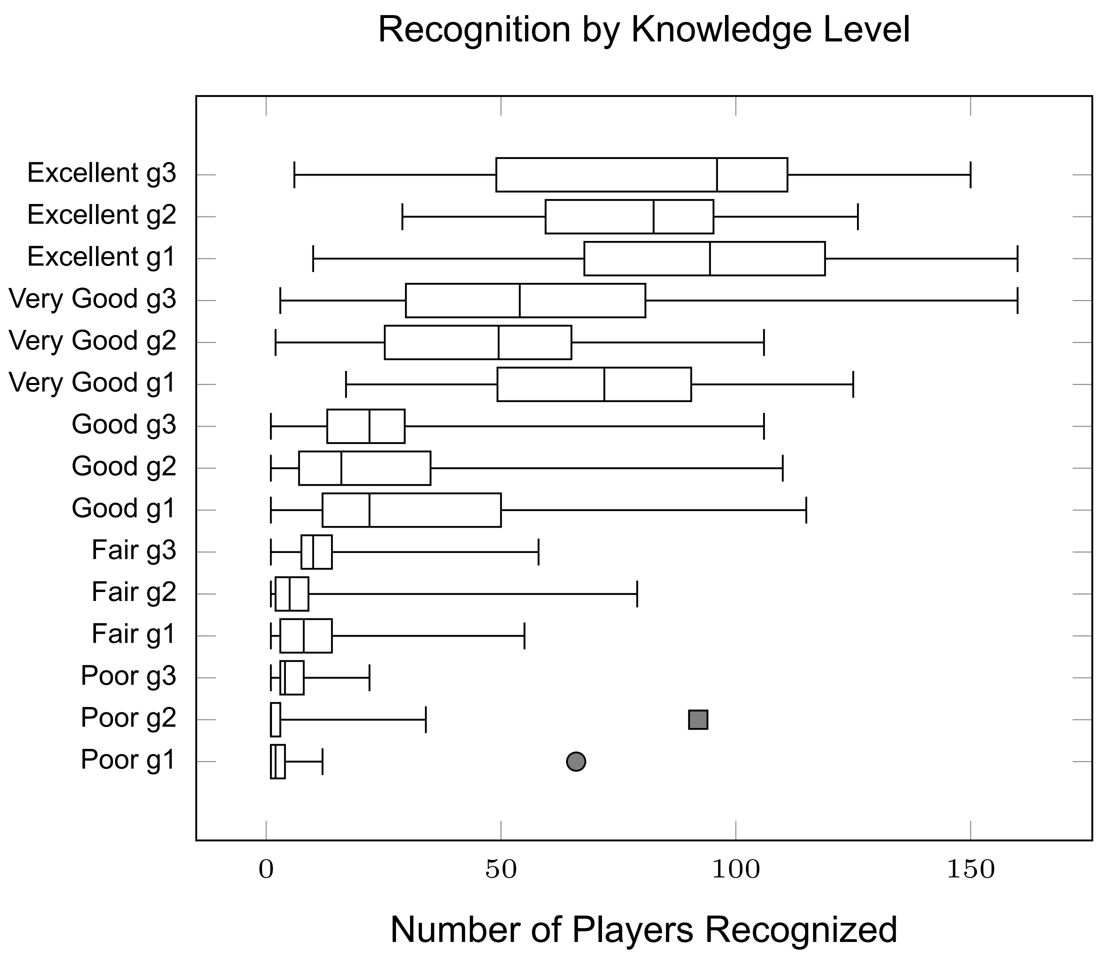

Figure 4: The number of players recognized by each self-reported knowledge level are shown in this figure. Respondents chose on a 5-point likert scale with poor representing the least knowledge and excellent representing the most. These values are in increasing order from bottom to top. They are represented further by each of the tree sections of the survey denoted g1, g2, g3, for groups one two and three.

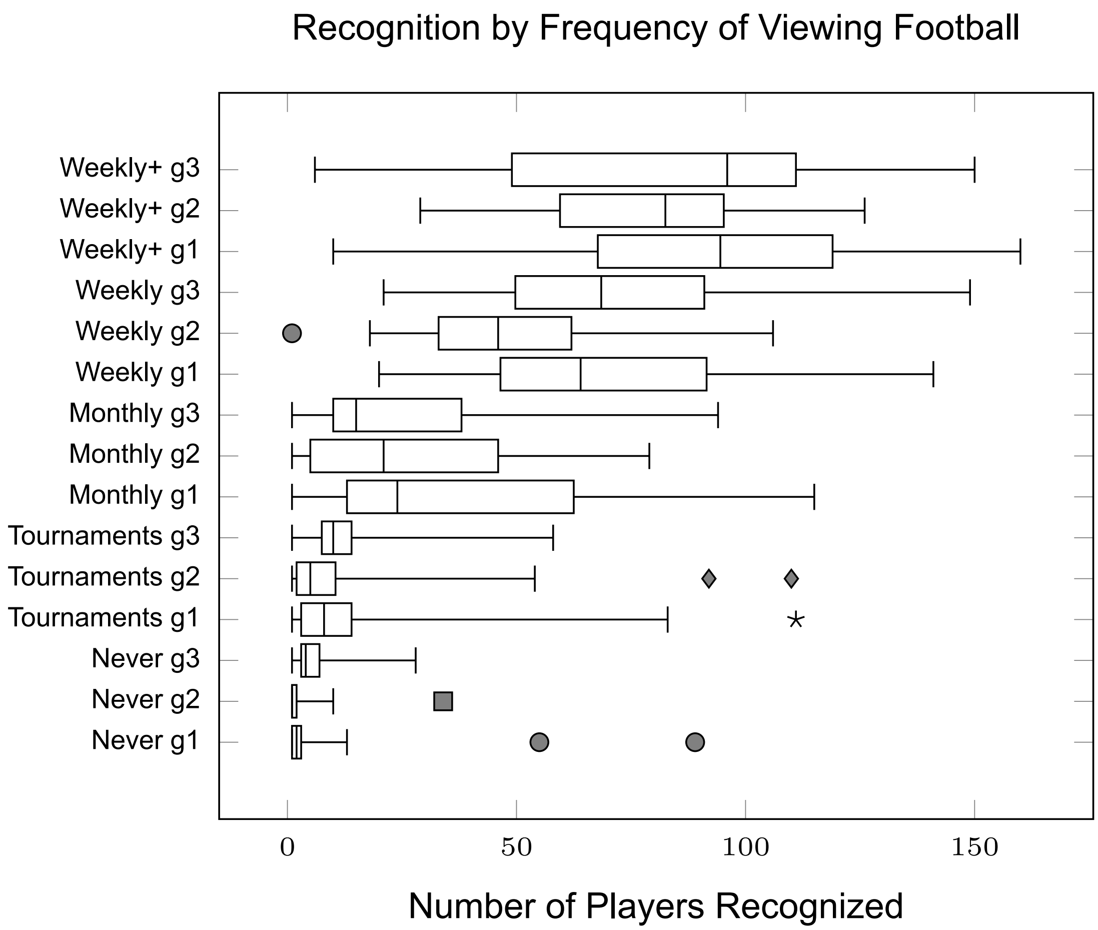

Figure 5: The number of players recognized is shown by the self-reported knowledge level of participants. Respondants chose on a 5-point likert scale for the option that best fit their football viewing habits. The options were: never, during major tournaments, monthly, weekly, and more than once a week. These are presented in increasing values from bottom to top in the figure. Each of the five values for frequency is present in the survey group, represented by g1, g2, and g2 for groups 1, 2, and 3.

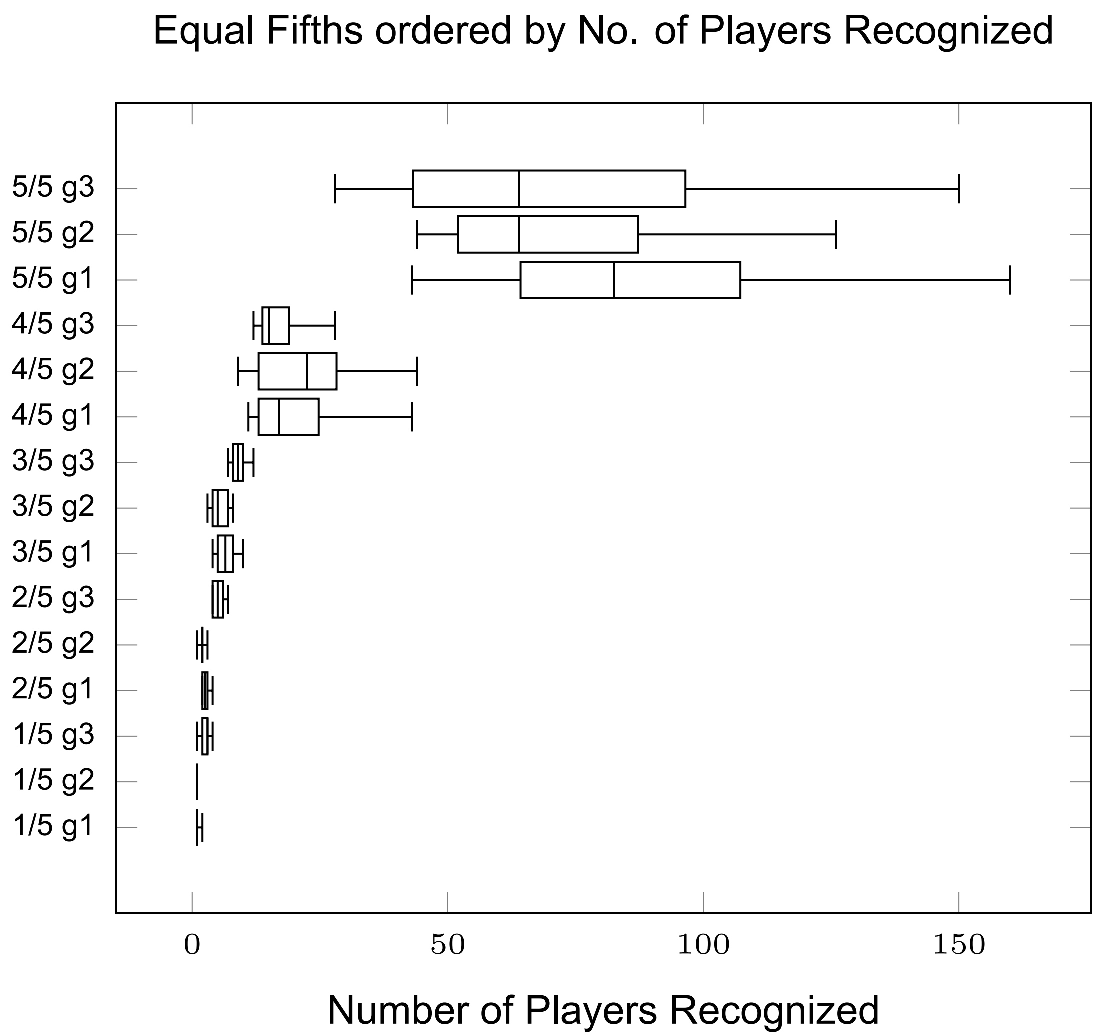

Figure 6: The participants were sorted by the number of players that they recognized. Then they were broken into five equal sized groups in order to correspond with the five groups of the tests based on self-reported knowledge and frequency of viewership. Each of the three survey groups is represented by g1, g2, and g3.

is a negative trend between level of knowledge and percent of matches accurately predicted. In all three instances, the team recognition model did little better, if not worse, than chance.

The lowest performance for the team recognition model was for those who reported viewing football more than once a week. In this group, the set with all participants predicted 38.5% of matches correctly while the set without German and Austrian respondents predicted 41% of matches correctly. These same low figures are found in the middle fifth of the trial sorted by number of players recognized. In this, 41% of matches were predicted correctly by the section with all respondents while the section without Austrian and German responses for their teams predicted 38.5% of matches correctly. The highest portion of matches correctly predicted was by participants who self-reported as having excellent knowledge of football in the section without Austrian and German responses for their teams. In this section, 59% of matches were correctly predicted, while the section with all responses for the same knowledge level was slightly lower at 56.4%.

There is little difference between the team based predictions of all participants and those without German and Austrian participants answering about Austrian and German teams. This makes sense, as whether one is German or Austrian is not likely to change if they have heard of each of these national teams. This is especially true since Germany won the World Cup in 2014 and the survey was conducted in Austria where there was discussion about the Austrian national team prior to the start of the

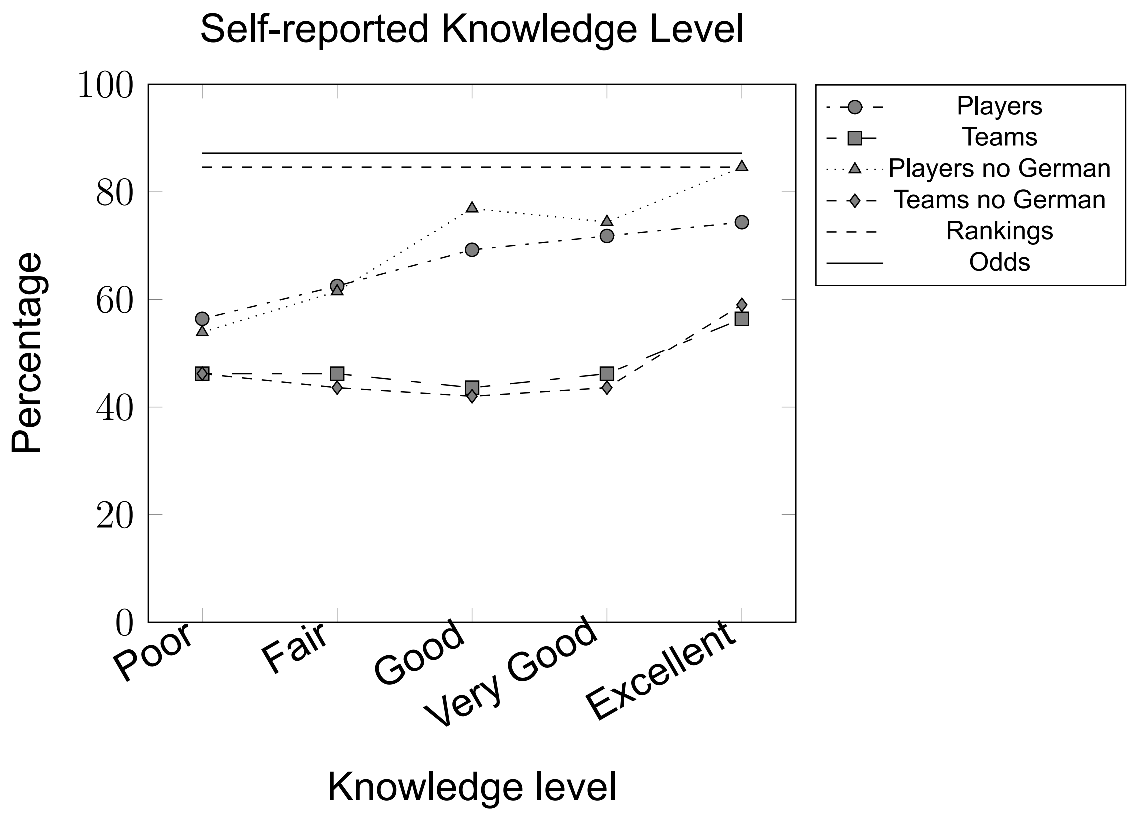

Figure 7: Triangles are for player-recognition without German or Austrian responses for their teams. Circles are for player-recognition for the whole sample. Squares are for team-recognition and diamonds are for team-recognition without German or Austrian responses for their teams.

tournament.

The atom-recognition model outperformed the team recognition model in nearly all instances. Only in the two samples of equal fifths model where groups were organized based off of number of players recognized was the atom-recognition outperformed by the team recognition model. This can be attributed to the fact that so few players were recognized on average

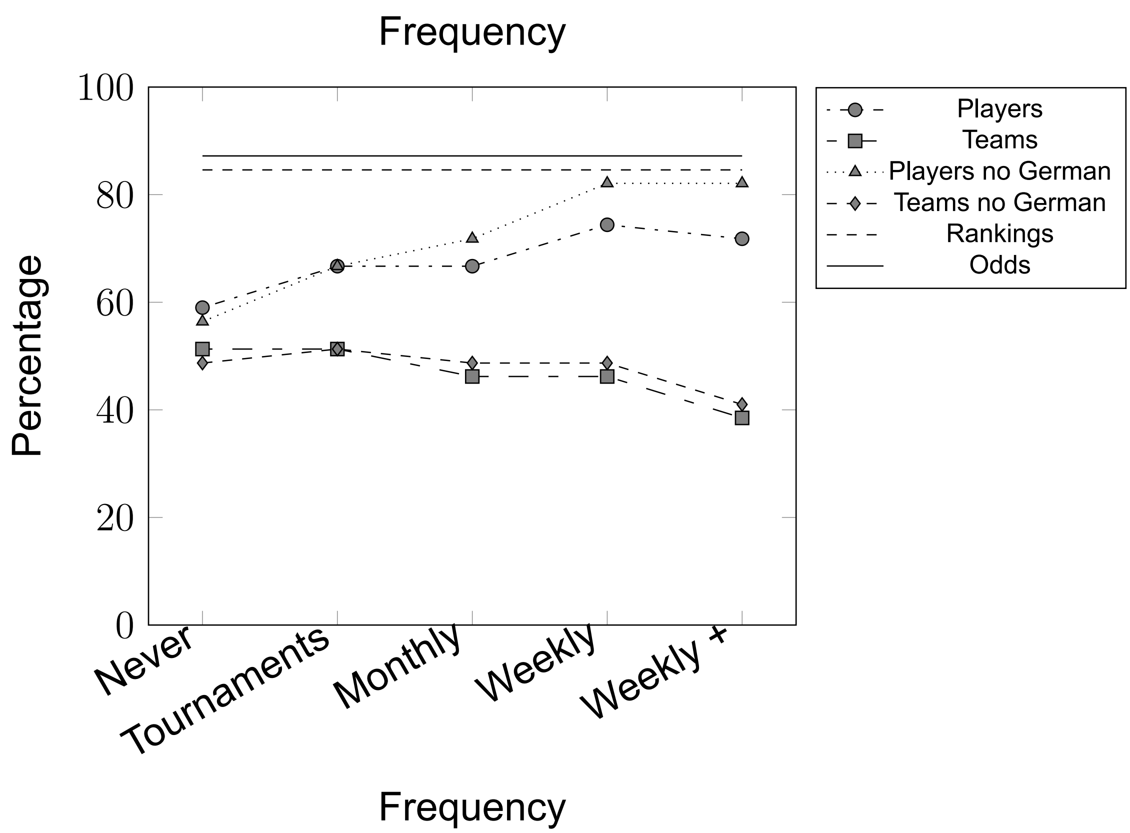

Figure 8: Triangles are for player-recognition without German or Austrian responses for their teams. Circles are for player-recognition for the whole sample. Squares are for team-recognition and diamonds are for team-recognition without German or Austrian responses for their teams.

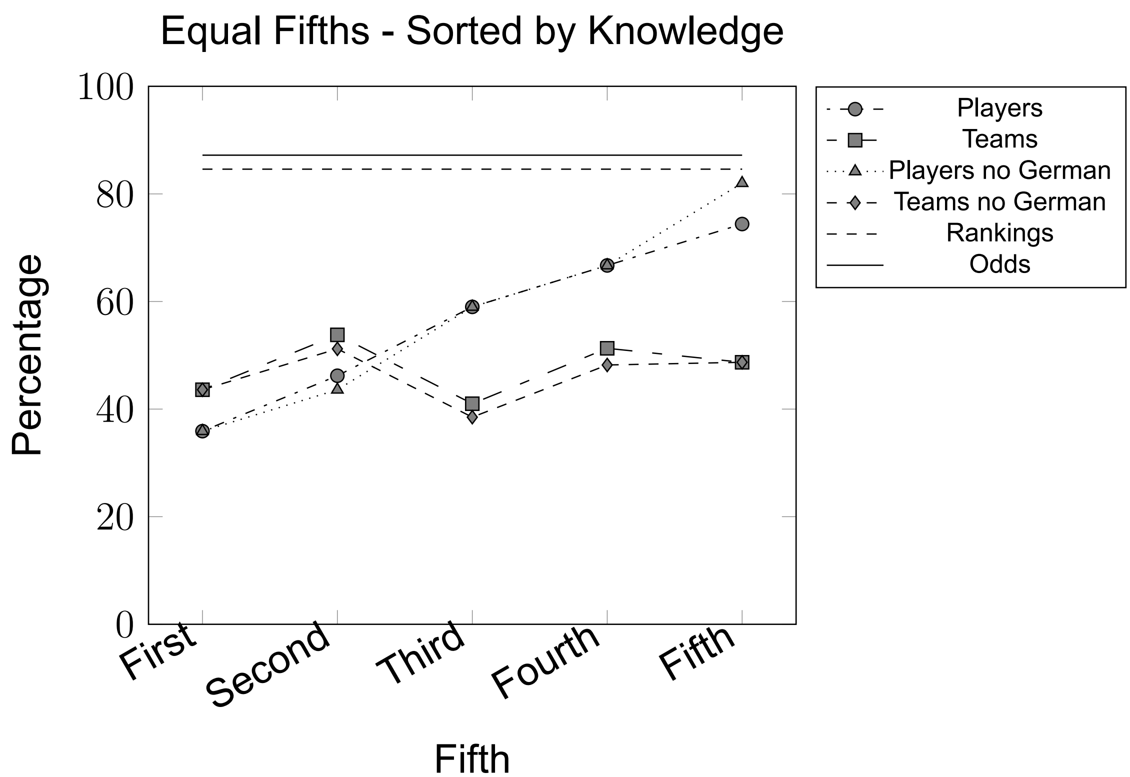

Figure 9: Triangles are for player-recognition without German or Austrian responses for their teams. Circles are for player-recognition for the whole sample. Squares are for team-recognition and diamonds are for team-recognition without German or Austrian responses for their teams.

	Group 1	G1 noGer	Group 2	G2 noGer	Group3	G3 noGer
Poor	123	25	80	14	107	18
Fair	74	20	69	14	79	15
Good	41	9	39	11	27	9
Very Good	26	3	30	6	29	8
Excellent	26	3	30	6	29	8

Table 1: Then sample size, N, for each knowledge level and each of the three survey groups. The sections labelled ”noGer” correspond to the sample size of all non German or Austrian participants. This is used only to forecast for the German and Austrian teams.

by respondents in the lowest recognition groups. In the first group, the most players recognized is four and in the second the most is seven. It is to be expected that any instance in which such a low portion of items are recognized would produce less than desired results. This is borne out in the prediction rate for these two instances being significantly below that of chance.

In all three analysis there is an upward trend to the percentage of games correctly foretasted as knowledge increases. This demonstrates that as self-reported knowledge of football, exposure to football matches, and number of players recognized increases the prediction accuracy of the models improve. This is evidenced by the fact that the lowest score for each of the three trials occurs at the lowest level of knowledge, viewership rate, or fifth respectively; the highest value thus is reached at either the highest or second highest knowledge group for each.

As previously mentioned, the lowest rate of correct forecasts was found in the analysis based on fifths of participants sorted by number of players.

	Group 1	G1 noGer	Group 2	G2 noGer	Group 3	G3 noGer
Never	101	20	54	7	74	8
Tournaments	115	31	117	26	136	36
Monthly	27	5	23	4	13	4
Weekly	23	3	18	3	16	1
Weekly +	24	4	27	10	20	5

Table 2: Then sample size, N, for each frequency of viewing football and each of the three survey groups. The sections labelled ”noGer” correspond to the sample size of all non German or Austrian participants. This is used only to forecast for the German and Austrian teams.

	Group 1	G1 no Germ	Group 2	G2 no Germ	Group 3	G3 no Germ
1 of 5	58	14	48	5	52	14
2 of 5	58	10	48	5	52	7
3 of 5	58	17	47	10	51	7
4 of 5	58	17	48	17	52	17
5 of 5	58	5	48	12	52	12

Table 3: Then sample size, N, for each fifth of respondents and each of the three survey groups. The section ”1 of 5” represents participants who recognized the fewest football players while the section ”5 of 5” represents the section with the most recognized players. The sections labelled ”no Germ” correspond to the sample size of all non German or Austrian participants. This is used only to forecast for the German and Austrian teams.

	Players	P no Germ	Teams	T no Germ	Odds	Ranking
Poor	56.41%	53.85%	46.15%	46.15%	87.18%	84.62%
Fair	61.54%	61.54%	46.15%	43.59%	87.18%	84.62%
Good	69.23%	76.92%	43.59%	41.03%	87.18%	84.62%
Very Good	71.79%	74.36%	46.15%	43.59%	87.18%	84.62%
Excellent	74.36%	84.62%	56.41%	58.97%	87.18%	84.62%

Table 4: The percent of matches which were correctly predicted by each of the self-reported frequencies of viewing football. The sections labelled ”no Germ” correspond to the sample in which only non German or Austrian participants were used to rate the players and team of Austria and Germany.

	Players	P no Germ	Teams	T no Germ	Odds	Ranking
Never	58.97%	56.41%	51.28%	48.72%	87.18%	84.62%
Tournaments	66.67%	66.67%	51.28%	51.28%	87.18%	84.62%
Monthly	66.67%	71.79%	46.15%	48.72%	87.18%	84.62%
Weekly	74.36%	82.05%	46.15%	48.72%	87.18%	84.62%
Weekly +	71.79v	82.05	38.46	41.03	87.18	84.62

Table 5: The percent of matches which were correctly predicted by each of the self-reported knowledge levels of football. The sections labelled ”no Germ” correspond to the sample in which only non German or Austrian participants were used to rate the players and team of Austria and Germany.

	Players	P no Germ	Teams	T no Germ	Odds	Ranking
1 of 5	35.90%	35.90%	43.59%	43.59%	87.18%	84.62%
2 of 5	46.15%	43.59%	53.85%	51.28%	87.18%	84.62%
3 of 5	58.97%	58.97%	41.03%	38.46%	87.18%	84.62%
4 of 5	66.67%	66.67%	51.28%	48.72%	87.18%	84.62%
5 of 5	74.36%	82.05%	48.72%	48.72%	87.18%	84.62%

Table 6: for each fifth of respondents and each of the three survey groups. The section ”1 of 5” represents participants who recognized the fewest football players while the section ”5 of 5” represents the section with the most recognized players. he sections labelled ”no Germ” correspond to the sample in which only non German or Austrian participants were used to rate the players and team of Austria and Germany.

For the lowest group, only 35.9% of matches were correctly predicted, this figure was the same for both the samples with all participants and that without German and Austrian participants answering about their teams. The highest prediction rate was found in the self-reported knowledge trial for those reporting excellent knowledge of football. The part without German and Austrian responses for their countries achieved a forecasting accuracy rate of 84.6%. This is equal to the rate predicted by the FIFA rankings. Additionally, each of the top sections in the other trials also found results which are only roughly 2% off from this high mark.

Looking at the effect that excluding Germans and Austrians from answering about their teams has, we can see that it does indeed have an effect. Interestingly though, this appears to only be the case at the higher knowledge levels of each trial. For both the trials based off of self-reported knowledge and frequency of viewership, the ”no German” section starts to outperform the section with all respondents. For the trial based on equal fifths, this point of differentiation only happens at the uppermost group, with all previous groups being identical or nearly identical in outcome.

7 Discussion

The mechanism of excluding Germans and Austrians did work in the intended way. By looking at the rankings for average recognition of the teams players and the teams crest one can see a difference in the values for German and Austria. With the German and Austrian answers included, the scores for the two countries increase significantly for the player-recognition score. For the team-crest recognition values, there was a nominal difference between the two. This is borne out by the fact that there is little difference between the two groups in all three trials. The German team crest had nearly universal recognition by both groups, while the Austrian crest enjoyed high recognition as well despite the design change for the tournament.

When looking at the results on the basis of individual games, one thing that was clear was that even with controlling for Austrian and Germans, these teams still seemed to be overestimated in the models. Outside of the games which were upsets, the other games which were incorrectly predicted were ones in which the model put either Germany or Austria to win over a team that they lost to. This demonstrates that the control measures did have some impact, as can be seen in its better performance, it is evident that there was greater knowledge about the Austrian and German teams among participants who are not from the two countries.

This gets at the tricky issue of citizenship. I used the citizenship that people most identify as in order to control for German and Austrian knowledge. This is an imperfect means of ascertaining if participants had specific culture-specific knowledge. For example, there could be participants who have grown up in Austria but do not have an Austrian passport. This would lead them to check the country of their passport, while for all intents and purposes they would be just as Austrian as any Austrian citizen. Additionally there may be those who have dual citizenship from birth who pick one of the two options, leaving the other un-signaled but still having influence on their decision making.

One potential issue in the team crest recognition part of the survey was how participants interpreted the word ”recognize”. Many of the team crests have the name of the team or country as a part of the design of the crest. These are done in the local language which was not a problem since most participants can’t read Danish or Finnish. Some of the teams which have English as the local language had English writing as a part of their crest and this proved more problematic. Especially in the case of Scotland which has the country name written on it. This resulted in Scotland receiving much higher than anticipated recognition values for the team crest than would have been expected given the stature of the team. After all the theory is that exposure through the media and people wearing the jersey in public would lead to recognition. However I think that outside of dedicated fans, not many people are wearing Scotland jerseys in the streets of Vienna.

This suggests that participants were interpreting ”do you recognize this team crest?” as ”do you know what team this is for?”. This is problematic because different countries crests have different levels of ”guessability”. That is, it would be easier to guess that the Swiss National Team’s crest belongs to them as it it simply their flag (square and all) whereas other countries are more difficult. Hungary for example is the Hungarian coat of arms with the Hungarian crown sitting on top. Given that Scotland seems to have been guessed at, it is likely that others would also have been guessed at by some of the participants.

Despite these issues with some of the responses the results are still quite promising, especially compared to previous research in sports forecasting. There is no way to directly compare odds forecasting across tournaments because of stochastic element of any sports outcome. In order to compare across studies I will use the betting odds in each study as the reference, since this is seen as the gold standard of sports forecasting. Then the percentage of the betting odds which is correctly predicted by the model in each paper will be used to compare across papers and tournaments. This way there is a much more level playing field than simply comparing model percentages across tournaments. Then when there are multiple tests in a paper I will take the highest percent that they were able to achieve.

Table 7: Performance of different Sports Recognition Heuristic Studies

1greywhite

Tournament	**Games**	**Recognition Heuristic**	**Odds**	**Percent of Odds**
Euro 2020	39	84.6%	87.2%	97.1%
World Cup 2006	48	84%	95%	88.4%
Euro 2008	24	62%	64%	96.9%
Wimbledon 2005	127	68%	79%	86.1%
Wimbledon 2003	96	72%	79%	91.1%

Both of the tennis studies of Wimbledon by Serwe and Frings (2006) as well as Scheibenhenne and Bröder (2007) found that the more knowledgeable of the groups surveyed performed better. Both of these studies collected recognition data for the individual tennis players and then created a ranking of the players by their recognition rate. This is an approach which is essentially analogous to the player atom-recognition carried out for the football teams because it looks at the finest level of analysis. This allows participants to apply a greater portion of their knowledge through recognition cues.

On the other hand, analysis like that of Ayton et al. (2011) when comparing the effectiveness of Turkish and English students' ability to forecast English football matches was conducted at the team level. They asked the participants to indicate which team they thought would win the match and then collected 7-point familiarity data on the teams. With the familiarity data for the pairs of teams they were then able to determine, based off of the difference in the two scores if choices were made using the recognition heuristic. Since the English participants are not being asked questions which allow them to make greater use of the knowledge which they have, they are hamstrung by only being able to signal a smaller portion of their knowledge through the team-level analysis.

This can also be seen in the work by Goldstein and Gigerenzer (2002) which the first instance of the less is more effect was put forth. In their study, they presented US Americans and Germans with random pairs of the largest cities in Germany. They found that the participants from the US did as well in guessing which of the two German cities was larger as the German participants. This though again is likely due to the coarseness of the recognition information that was gathered. Gathering recognition data on the level of cities would be analogous to gathering recognition data on the level of football teams as a whole. While this may work, it is likely that there is a better way forecast the item in question.

An example of the difference in coarseness can be seen in the results of this paper; particularly the difference between the team-level recognition forecasts and the player-level recognition forecasts. For the team recognition results there is no correlation across the three tests between forecasting percentage and knowledge level. That is, as the participants know more about football their performance in the team-level analysis did not improve. This is contrasted with the player-level analysis in which there is a correlation between the football knowledge-level of the participants and their performance in the forecasting. In all three tests there is a positive correlation between knowledge level and prediction performance.

I argue that this is because the player-level analysis is a finer level of analysis which permits the greater knowledge to be put to use, where coarser levels of analysis do not permit this. In the case of Ayton et al. (2011) it seems likely that if the participants were asked for recognition data for the players on the teams in the FA Cup ties that the English participants would perform better than the Turkish participants. Not only this, but that the forecasting accuracy rate would be increased for the groups with greater knowledge. Although it could also be the case that the groups with less knowledge perform better at a finer level of analysis.

If groups with less knowledge are performing as well in forecasting tasks as groups with more knowledge, this certainly is a remarkable outcome as noted by Goldstein and Gigerenzer. If the goal is simply to forecast well with as little information as possible then one can stop there. However, if one is to try to achieve predictions that are as accurate as possible, models have to be constructed which are able to take advantage of all the available information that participants have. In order to do this, the finest level of analysis should be used and participants with a higher level of knowledge should be sought out. In this paper I found that prediction performance increased as the knowledge of the participants increased. This trend continued upwards throughout the data and no decrease was found. This suggests that yet further increases in knowledge may lead to further increases in forecasting accuracy. Further research should focus on finding at what point the accuracy starts to decline again. There must be such a point because respondents who for example know every professional footballer in Europe then would select every player in the questionnaire. This would lead to a team recognition rate or 100% for every team. Since every team has the same value, no teams would be predicted ”over” another and thus the model would have a prediction rate of 0%.

8 Conclusion

The recognition heuristic has shown itself to be adept at providing surprisingly accurate forecasts despite limited information. Additionally, forecasting using the recognition heuristic is not as prone to over fitting as traditional regression modelling. The combination of both of these factors makes forecasting using the recognition heuristic a valid option in many cases, both in terms of its accuracy and its efficiency. Much progress has been made in the field since the discovery of the resignation heuristic; there is now a good understanding of some of the fields in which it is applicable. Further research should focus on looking to establish a set of best practices for making predictions using the recognition heuristic. Through establishing these best practices, researchers will be able to produce more accurate forecasts with less effort.

A good deal of research has been conducted in the field of sports forecasting due to the general level of knowledge among the populous, clear outcomes, and comparability to other means of forecasting. The recognition heuristic is merely one of an adaptive toolbox of fast and frugal decision making strategies which enable us to make accurate decisions in the face of limited information. They should not be shied away from because of their rudimentary nature, rather, leaned into in cases where they provide good results. This would provide significant time savings for many decisions without resulting in a significant reduction in accuracy.

9 Thanks

I would like to thank Jana May Purington for her feedback as well as Alice Lee and Kira Vasiliev for their help with LaTeX.

References

Andersson, P., Edman, J., & Ekman, M. (2005). Predicting the world cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting*, *21*(3), 565–576.

Andersson, P., Memmert, D., & Popowicz, E. (2009). Forecasting outcomes of the world cup 2006 in football: Performance and confidence of bettors and laypeople. *Psychology of Sport and Exercise*, *10*(1), 116–123.

Andersson, P., & Rakow, T. (2007). Now you see it now you don't: The effectiveness of the recognition heuristic for selecting stock. *Judgment and Decision Making*, (2), 29–39.

Ayton, P., Önkal, D., & McReynolds, L. (2011). Effects of ignorance and information on judgments and decisions. *Judgment and Decision Making*, *6*(5), 381–391.

Boulier, B. L., & Stekler, H. O. (1999). Are sports seedings good predictors?: An evaluation. *International Journal of Forecasting*, *15*(1), 83–91.

Boulier, B. L., & Stekler, H. O. (2003). Predicting the outcomes of national football league games. *International Journal of forecasting*, *19*(2), 257–270.

Bröder, A., & Eichler, A. (2006). The use of recognition information and additional cues in inferences from memory. *Acta psychologica*, *121*(3), 275–284.

Brown, A., & Reade, J. J. (2019). The wisdom of amateur crowds: Evidence from an online community of sports tipsters. *European Journal of Operational Research*, *272*(3), 1073–1081.

Buraimo, B., & Simmons, R. (2008). Do sports fans really value uncertainty of outcome? evidence from the english premier league. *International Journal of Sport Finance*, *3*(3).

Christenson, M. (2010). Psychic octopus paul predicts spain to beat holland in world cup final. https://www.theguardian.com/football/2010/jul/09/psychic-octopus-paul-picks-spain

Chun, S., Gentry, J. W., & McGinnis, L. P. (2005). Ritual aspects of sports consumption: How do sports fans become ritualized? *ACR Asia-Pacific Advances*.

Clarke, S. R., & Dyte, D. (2000). Using official ratings to simulate major tennis tournaments. *International transactions in operational research*, *7*(6), 585–594.

Del Corral, J., & Prieto-Rodríguez, J. (2010). Are differences in ranks good predictors for grand slam tennis matches? *International Journal of Forecasting*, *26*(3), 551–563.

FIFA. (n.d.-a). https://resources.fifa.com/image/upload/fifa-world-ranking-technical-explanation-revision.pdf?cloudid=edbm045h0udbwkqew35a&_branch_match_id=link-596649936691732161

FIFA. (n.d.-b). More than half the world watched record-breaking 2018 world cup https://www.fifa.com/tournaments/mens/worldcup/2018russia/media-releases/more-than-half-the-world-watched-record-breaking-2018-world-cup.

Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of english football. *International journal of forecasting*, *21*(3), 551–564.

Franck, E., Verbeek, E., & Nüesch, S. (2010). Prediction accuracy of different market structures—bookmakers versus a betting exchange. *International Journal of Forecasting*, *26*(3), 448–459.

Gaissmaier, W., & Marewski, J. N. (2011). Forecasting elections with mere recognition from small, lousy samples: A comparison of collective recognition, wisdom of crowds, and representative polls. *Judgment and Decision Making*, *6*(1), 73–88.

Galton, F. (1907). Vox populi (the wisdom of crowds). *Nature*, *75*(7), 450–451.

Gigerenzer, G., & Engel, C. (2006). *Heuristics and the law*. MIT press.

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making*, *6*(1), 100–121.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A brunswikian theory of confidence. *Psychological review*, *98*(4), 506.

Gigerenzer, G., & Todd, t. A. R. G., Peter M. (1999). *Simple heuristics that make us smart*. Oxford University Press, USA.

Gil, R. G. R., & Levitt, S. D. (2007). Testing the efficiency of markets in the 2002 world cup. *The Journal of Prediction Markets*, *1*(3), 255–270.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological review*, *109*(1), 75.

Hafenbrädl, S., Waeger, D., Marewski, J. N., & Gigerenzer, G. (2016). Applied decision making with fast-and-frugal heuristics. *Journal of Applied Research in Memory and Cognition*, *5*(2), 215–231.

Herzog, S. M., & Hertwig, R. (2011). The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. *Judgment and Decision making*, *6*(1), 58–72.

Hoffrage, U. (2011). Recognition judgments and the performance of the recognition heuristic depend on the size of the reference class. *Judgment and Decision Making*, *6*(1), 43–57.

Jordet, G., Hartman, E., Visscher, C., & Lemmink, K. A. (2007). Kicks from the penalty mark in soccer: The roles of stress, skill, and fatigue for kick outcomes. *Journal of Sports Sciences*, *25*(2), 121–129.

Lebovic, J. H., & Sigelman, L. (2001). The forecasting accuracy and determinants of football rankings. *International Journal of Forecasting*, *17*(1), 105–120.

Leitner, C., Zeileis, A., & Hornik, K. (2010). Forecasting sports tournaments by ratings of (prob) abilities: A comparison for the euro 2008. *International Journal of Forecasting*, *26*(3), 471–481.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of economic perspectives*, *17*(1), 59–82.

Marewski, J. N., Gaissmaier, W., & Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cognitive processing*, *11*(2), 103–121.

Newell, B. R., & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 923.

Odds portal - betting odds monitoring service. (n.d.). https://www.oddsportal.com/

Oppenheimer, D. M. (2003). Not so fast!(and not so frugal!): Rethinking the recognition heuristic. *Cognition*, *90*(1), B1–B9.

Pachur, T., & Biele, G. (2007). Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events. *Acta Psychologica*, *125*(1), 99–116.

Pawlowski, T., & Anders, C. (2012). Stadium attendance in german professional football–the (un) importance of uncertainty of outcome reconsidered. *Applied Economics Letters*, *19*(16), 1553–1556.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological review*, *109*(3), 472.

Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, *19*(3), 251–271.

Scheibehenne, B., & Bröder, A. (2007). Predicting wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting*, *23*(3), 415–426.

Serwe, S., & Frings, C. (2006). Who will win wimbledon? the recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making*, *19*(4), 321–332.

Smithson, M. (2010). When less is more in the recognition heuristic. *Judgment and Decision Making*, *5*(4), 230.

Snook, B., & Cullen, R. M. (2006). Recognizing national hockey league greatness with an ignorance-based heuristic. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *60*(1), 33.

Suzuki, K., & Ohmori, K. (2008). Effectiveness of fifa/coca-cola world ranking in predicting the results of fifa world cup finals. *Football Science*, *5*, 18–25.

Thaler, R. H., & Ziemba, W. T. (1988). Anomalies: Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic persp ectives*, *2*(2), 161–174.

TRTWorld. (2018). Meet the animals predicting winners of world cup in russia. https://www.trtworld.com/sport/meet-the-animals-predicting-winners-of-world-cup-in-russia-18232

UEFA.com. (2021). All the euro 2020 squads. https://www.uefa.com/uefaeuro-2020/news/0269-12464679fb2e-84fa175e53ca-1000--all-the-euro-2020-squads/

Vlastakis, N., Dotsis, G., & Markellos, R. N. (2009). How efficient is the european football betting market? evidence from arbitrage and trading strategies. *Journal of Forecasting*, *28*(5), 426–444.

von Martin Mohr, V., & 17.02.2022. (2022). Österreich - ausländer staatsangehörigkeiten 2022. https://de.statista.com/statistik/daten/studie/293019/umfrage/auslaender-in-oesterreich-nach-staatsangehoerigkeit/

Wegwarth, O., Gaissmaier, W., & Gigerenzer, G. (2009). Smart strategies for doctors and doctors-in-training: Heuristics in medicine. *Medical education*, *43*(8), 721–728.

Wübben, M., & Wangenheim, F. v. (2008). Instant customer base analysis: Managerial heuristics often "get it right". *Journal of Marketing*, *72*(3), 82–93.

Wunderlich, F., Berge, F., Memmert, D., & Rein, R. (2020). Almost a lottery: The influence of team strength on success in penalty shootouts. *International Journal of Performance Analysis in Sport*, *20*(5), 857–869.