Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 23 (2007) 415–426
www.elsevier.com/locate/ijforecast

Predicting Wimbledon 2005 tennis results by mere player name recognition

Benjamin Scheibehenne [a,*], Arndt Bröder [b]

[a] Max Planck Institute for Human Development, Lentzeallee 94, D-14195 Berlin, Germany
[b] University of Bonn and Max Planck Institute for Research on Collective Goods, Bonn, Germany

Abstract

The outcomes of matches in the 2005 Wimbledon Gentlemen's tennis competition were predicted by mere player name recognition. In a field study, amateur tennis players ($n=79$) and laypeople ($n=105$) indicated players' names they recognized, and predicted match outcomes. Predictions based on recognition rankings aggregated over all participants correctly predicted 70% of all matches. These recognition predictions were equal to or better than predictions based on official ATP rankings and the seedings of Wimbledon experts, while online betting odds led to more accurate forecasts. When applicable, individual amateurs and laypeople made accurate predictions by relying on individual name recognition. However, for cases in which individuals did not recognize either of the two players, their average prediction accuracy across all matches was low. The study shows that simple heuristics that rely on a few valid cues can lead to highly accurate forecasts.

Keywords: Sports forecasting; Simple heuristics; Recognition; Betting odds

1. Introduction

Common sense suggests that the prediction of event outcomes and other probabilistic inferences improve when more information is integrated. In contrast to statistical approaches, which try to integrate all potentially relevant information, a recent study by Andersson, Edman, and Ekman (2005) reports that American and Swedish students with little knowledge about soccer were nevertheless more successful in predicting the results of the Soccer World Cup 2002 than were soccer experts (sport journalists, soccer fans, and soccer coaches). Andersson et al.'s results are in line with a growing body of evidence showing that simple models, which only use minimal amounts of information, are often as accurate as complex statistical models that integrate many pieces of information (Gigerenzer, Todd, & The ABC Research Group, 1999). Along the same lines, Spyros Makridakis and his colleagues (Fildes & Makridakis, 1995; Makridakis & Hibon, 1979, 2000; Makridakis et al., 1982, 1993) showed in a series of studies (the so-called M-competitions) that for predicting real-life time series, simple models lead to more

* Corresponding author. Tel.: +49 30 8240 697; fax: +49 30 824 06 394.
E-mail address: scheibehenne@mpib-berlin.mpg.de (B. Scheibehenne).

0169-2070/$ - see front matter © 2007 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2007.05.006

accurate forecasts than statistically sophisticated methods. The same case was argued by Robyn Dawes (1979), who showed that in many situations an "improper" linear model that uses equal weights leads to better predictions than a model based on presumably "optimal" or fitted weights.

In this paper, we investigate the success of a specific decision rule, the fast and frugal recognition heuristic (Goldstein & Gigerenzer, 2002), in predicting the 2005 Wimbledon tennis tournaments. In doing so, we also provide a potential explanation as to why and how the recognition heuristic works in predicting the outcomes of sport events. Our work confirms and extends the results of Serwe and Frings (2006), who used a similar approach to study the 2003 Wimbledon tournaments.

First, we will briefly characterize the simple heuristics view and the recognition heuristic. Then we will describe alternative predictors of tennis success, before we present our own study and results in more detail.

1.1. The heuristics view

Past research on fast and frugal heuristics illustrates that, contrary to conventional wisdom, the reduction of model complexity and the amount of information utilized need not be a disadvantage (Gigerenzer et al., 1999; Gigerenzer, 2004). For example, it has been demonstrated that working memory limitations may be beneficial for children's acquisition of their first language (Elman, 1993), or for detecting correlations within real world environments (Kareev, 1995, 2000). Also, as has been shown by computer simulations, relying on one good reason instead of a multiple regression analysis with many variables can lead to better predictions of demographic and economic variables such as homelessness and school drop-out rates (Czerlinski, Gigerenzer, & Goldstein, 1999). Despite their low computational complexity, simple models such as heuristics can outperform complex forecasting models because they have very few free parameters, and only rely on a few, albeit important, pieces of information. Because of this, heuristics ignore the noise inherent in many data sets, and are thus more robust to overfitting, a problem that impairs the predictive power of many complex and highly parameterized models (Forster & Sober, 1994; Hertwig & Todd, 2003). This notion also fits with the finding that in a wide range of domains, the judgments of experts, as compared to non-experts, can be described by the use of fewer, but more important cues (Shanteau, 1992).

Of course, these findings do not imply that fast and frugal heuristics are foolproof, or that limitations are always good. Rather, the lesson learned from this line of research is that of what is known as ecological rationality: a prediction strategy is ecologically rational if it is able to exploit the structure of the domain in which it operates. Accordingly, in some rare cases, and in artificial experimental situations, it has been shown that the use of simple heuristics may also lead to systematic judgmental errors sometimes referred to as 'biases' (Kahnemann & Tversky, 1996, 2000; but see Gigerenzer, 1996, for a criticism of this perspective). Yet past research has shown in many situations that people are often well adapted to their environment such that they select the strategies or heuristics that work best in a given situation (Rieskamp & Otto, 2006). In summary, the question of whether a fast and frugal heuristic does well in predicting an outcome is as much a question of the properties of the heuristic as of the structure of the domain to which it is applied. This relationship has already been emphasized by the Nobel laureate Herbert Simon, who stated that "Human rational behavior is shaped by a pair of scissors whose two blades are the structure of task environments and the computational capabilities of the actor" (Simon, 1990, p. 7).

1.2. The recognition heuristic

One particularly simple and non-compensatory heuristic is the recognition heuristic, which works as follows: if you have to judge which one of two objects scores higher on a criterion, and you recognize one of them but not the other, then choose the recognized object (Gigerenzer & Goldstein, 1996). Although it is extremely simple, this strategy can be very successful if the probability of recognition is highly correlated with the criterion to be judged. Hence, there are many domains in which mere recognition has been shown to be a valid predictor: for example, in judging city sizes (Goldstein & Gigerenzer, 2002), predicting stock performance in a bull market (Borges, Goldstein, Ortmann, & Gigerenzer, 1999; but see Boyd, 2001, for a bear market), predicting record sales of pop stars (Herzog, 2005), the quality of American colleges

(Hertwig & Todd, 2003), and the results of political elections (Marewski, Gaissmaier, Dieckmann, Schooler, & Gigerenzer, 2005). It has also been shown to predict sports success. In a study on predicting the outcome of the F.A. cup, a major knockout tournament for English football clubs, Ayton and Önkal (2006) found that by relying on name recognition, Turkish students who knew little about the English league were almost as successful as British students in predicting the outcome of the games. Similar results were found by Pachur and Biele (2007) in predicting the 2004 European Soccer Championships, by Serwe and Frings (2006) in predicting the outcome of the 2003 Wimbledon tennis competition, and by Snook and Cullen (2006) in predicting which of two Canadian hockey players has more career points.

Although there is debate about the claim that recognition is the *only* cue used in probabilistic inferences when applicable (Bröder & Eichler, 2006; Newell & Fernandez, 2006; Newell & Shanks, 2004; Pohl, 2006; Richter & Späth, 2006), all these studies show that people's inferences are heavily influenced by recognition information. Thus, people apparently use the recognition cue in forecasts.

However, in a dynamic setting such as sports predictions, the success of the recognition heuristic might be impaired, because sporting excellence is relatively short-lived while name recognition remains relatively stable over time. For example, Björn Borg is still a well known player even though his professional career is long over. In their study on the prediction of the men's Wimbledon tournament 2003, Serwe and Frings asked laypeople and tennis amateurs which of the player names they knew. Based on this name recognition data, they predicted that each match would be won by the player whose name was recognized by more participants. What they found was that, despite the dynamic change in the criterion, recognition still served as a valid predictor in predicting the results of the Wimbledon tournament.

At the same time, if the tennis environment is indeed "dynamic", the positive results for the recognition heuristic in 2003 might have been a coincidence. For example, recognition of company names was a good predictor of stock performance in the study of Borges et al. (1999), but it completely failed in a different market situation (Boyd, 2001). Hence, before praising partial ignorance as the silver bullet in tennis prediction, one has to show that Serwe and Fring's results are systematic. Also, there is a need to better understand the underlying reason for why, and in which domains, recognition serves as a valid predictor. The present study aims to further explore these issues by testing the recognition heuristic in predicting the 2005 Wimbledon Gentlemen's singles competition.

2. Method

To test the success of the recognition heuristic in predicting the outcome of the 2005 Wimbledon competition, we asked laypersons, as well as amateur tennis players, to indicate which of the names of the 128 Wimbledon contestants (112 regular players plus 16 qualifiers) they recognized. From these judgments, predictions were generated for all 127 matches of the tournament, from the first round of 64 to the final. Predictions were generated in two ways: first, across all participants in our study, each Wimbledon player was assigned a rank indicating how many participants recognized him, and for each match, the winner was predicted by that ranking. Second, for each individual participant in our study, predictions were generated for the matches in which the recognition heuristic could be applied. This is the case when one of the players is recognized and the other is not. The former ranking method may be more effective, because partial ignorance is aggregated across participants. The latter individual method is psychologically more plausible because every participant, of course, has access only to his or her own partial knowledge.

Both prediction methods were compared to predictions based on the official ATP Champions Race ranking (ATP-CR) and the ATP Entry ranking (ATP-ER). The ATP-CR ranks players according to their performance during the calendar year. In January, each player starts with zero points, and the performance at major tournaments is summed up over the course of the year. Although this rarely happens, ties between two or more players may occur. At the end of the year, the ATP-CR determines the year-end world Number 1 player. The ATP-ER, commonly referred to as the 'world ranking', reflects players' performance at major tournaments during the immediate past 52 weeks at any point. A player who ranks first on this list is said to be the Number 1 player in the world. In the domain of sports forecasting, models that rely on publicly listed data such as past performance and team rankings have proven to be valid predictors that outperform newspaper tipsters (Forrest & Simmons, 2000). The predictive power of player rankings and expert seedings has also been shown for other sports like basketball, tennis, and (American) football (Boulier & Stekler 1999, 2003).

As additional standards of comparison, we also included betting odds taken from several international online betting sites and the Wimbledon seedings (SEED). The 32 seeded players are determined by an official committee of experts that evaluates players' ability to perform on the rye grass of the Wimbledon tennis courts. Players who are seeded do not play against each other until late in the tournament and the two players seeded first and second will not play against each other until the final. While the committee uses the official ATP rankings as an orientation, it maintains the prerogative to deviate from it. Accordingly, in the year of our study (2005), 4 of the top 32 ATP-ER were not seeded, and 7 of the top 32 ATP-CR were not seeded. In addition to the recognition data, we also asked our participants to predict the results of the individual matches in the Wimbledon competition.

2.1. Materials and procedure

As the pairings unfolded during the course of the tournament, we collected data at two points in time. Data on the prediction of the first round (64 games) were collected between the end of the qualifying round and the start of the first round. Data on the prediction of the fourth round (8 games) and the quarterfinals (4 games) were collected after the end of the third round and before the start of the fourth round. In both sampling waves, data were collected using a questionnaire that consisted of a recognition section and a prediction section. The recognition section of the questionnaire listed the full names of all 128 male players who qualified for the 2005 Wimbledon Gentlemen's singles competition. For each player, participants were asked whether they knew the name or not. Note that for both rounds this left only a narrow window of little more than 24 h to prepare the questionnaires and collect the data.

The prediction section that followed differed in the two sampling waves. For the first wave, the questionnaire listed the 64 fixtures of the first round. For the second wave, the 8 fixtures of the fourth round and the 4 fixtures of the quarterfinals were listed. Since the opponents of the quarterfinals were not determined at that point in time, all 16 possible combinations for the quarterfinal encounters were listed. For the analysis we only used the combinations that were actually played. In each case, participants had to tick the name of the player they predicted to win the game.

To distinguish laypersons from amateur players, participants were asked whether they were a member of a tennis club and whether they were active tennis players. To increase the chances of finding amateurs, half of the questionnaires were administered on the sites of several tennis clubs in Berlin, while the other half were administered in several public parks within Berlin. To control for order effects, two different questionnaire versions were used, with the players' names in randomized orders. Participants received €2.00 for filling out a questionnaire. In addition, participants earned a lottery ticket for each correct prediction. The number of correct predictions determined their chances of winning €50.00 at the end of the study.

In each sampling wave, the recognition section of the questionnaire preceded the prediction section. As one reviewer noted, this might have primed participants to make use of recognition in their predictions. While we cannot fully rule out this possibility, we believe that we counteracted this by setting a monetary incentive for making correct predictions. More importantly, we were concerned that by doing it the other way round (first prediction, then recognition) we might have induced recognition because the participants would have already had seen the names before, a problem that researchers have faced in the past (Marewski et al., 2005). As having reliable recognition data was paramount for our analyses, we decided to ask for it first.

Since the data were collected in the field, some questionnaires were not fully completed. We excluded every participant with 10 or more missing values in either the recognition or the prediction segment.

2.2. Differences and extensions to previous work by Serwe and Frings

Our methodology is similar to the one used by Serwe and Frings (2006), but it extends their approach in several important ways. For the first round of 64, Serwe and Frings asked their participants about name recognition, but did not ask them to make predictions. In our study, we elicited recognition and prediction judgments from the same participants for the first round, which allowed us to more reliably assess the prescriptive, as well as the descriptive, value of the recognition heuristic. Moreover, Serwe and Frings did not include the 16 players that qualified for the competition in the week prior to the start of the tournament, while our study included all 128 Wimbledon players (rather than only 112 as in their study). Serwe and Frings sampled amateurs from one specific tennis club and recruited university students as laypeople. To rule out the possibility that their results were due to special properties of this sample (e.g., members of one single tennis club might recognize the same player names) we used a sample of amateur tennis players from 9 different tennis clubs. We also used a more heterogeneous sample of laypeople by recruiting participants "on the street". To further analyze when and why the recognition heuristic works, we also collected data on the public media coverage of the players.

2.3. Participants

From the initial sample, 26 participants were excluded due to missing values, which resulted in a total of 184 remaining participants for both weeks; 105 of them were laypeople and 79 were members of a tennis club, henceforth called *amateurs*. 24% of the amateurs were male and 50% of the laypeople were male. The average age of the amateurs was 37 years (SD = 14.9), and that of the laypeople was 30 years (SD = 11.0).

3. Results

3.1. Prediction of recognition ranking

On average, laypeople recognized 11.1 of the 128 players, or 9%. Two thirds of the laypeople recognized between 2% and 11% of the players, and there were 0.2 missing values on average. Amateurs recognized 49.9 players, or 39%. Two thirds of the amateurs recognized between 13% and 53%, and on average there were 0.9 missing values. From these data, a ranking was calculated based on how often each of the 128 players was recognized by our participants. Under the assumption that recognition is not random, but is related to the criterion of interest (here, the tennis players' success), the recognition heuristic predicts that the player that is recognized by more participants is more likely to win a match. The rule derived for predicting the outcomes of the Wimbledon matches is straightforward: if one player is recognized by more people, predict that this player will win the game. If both players are recognized by the same number of people, guess. In total, 127 matches were played during the tournament that could be used to test this prediction. We calculated three recognition rankings, one for the laypersons (RR-Lay), one for the amateurs (RR-Amateurs), and one for all participants (RR-All). The amateur ranking and the laypeople ranking are highly correlated ($\varphi = .87$), and therefore the predictions based on the rankings are quite similar. RR-Lay led to correct predictions in 84 of the 127 games, and it had to guess the outcome of two matches; thus, it successfully predicted 67% of all of the games. RR-Amateurs and RR-All both had to guess in one case. RR-Amateurs successfully predicted 86 matches (68%), and RR-All predicted 89 matches (70%). But how good is it to get 70% correct predictions?

3.2. Comparison of the recognition ranking with other rankings

In order to assess the predictive power of our recognition rankings, we compared them to ATP-ER, ATP-CR, and SEED. Both ATP rankings were collected one day before the official start of the tournament. As the Wimbledon tournament takes place in the middle of the year, the ATP-CR and ATP-ER are highly correlated, and therefore often lead to similar predictions. Table 1 gives an overview of the correlations among all of the rankings.

The rule for predicting the matches from the ATP rankings and the seedings worked like the rule used for the recognition rankings: predict that the player with the higher ranking will win the game; if two players have the same ranking, guess.[1] Based on this rule, ATP-ER never had to guess, and correctly predicted 88 of the 127 games (69%). ATP-CR had to guess once and correctly predicted 89 games (70%). As only 32 players were seeded, SEED had to guess in 42 cases, and correctly predicted 68 games (54%). An analysis

[1] Guessing occurs mainly for the SEED ranking because the Wimbledon committee only seeds 32 players. To make use of the SEED ranking, we assigned a dummy-seed of 33 to all unseeded players.

Table 1
Spearman correlation between the rankings ($N=127$)

	Recognition ranking — laypeople	Recognition ranking — all participants	ATP champions race	ATP entry ranking	Wimbledon seedings
Recognition ranking — amateurs	.87	.98	.68	.69	.73
Recognition ranking — laypeople		.94	.58	.52	.68
Recognition ranking — all participants			.67	.64	.73
ATP champions race				.87	.88
ATP entry ranking					.86

of variance reveals no difference between the ranking predictions ($F[5,756]=0.13$). Thus, even though most participants in our study were far from being tennis experts, the predictions based on recognition rankings were as good as the predictions from both ATP rankings, and the predictions of the Wimbledon expert committee.

3.3. Predictions of betting odds

We also compared the predictions of the recognition rankings to the betting odds for each individual game from five different online bookmakers.[2] For all five online bookmakers, the predictions based on betting odds for all matches are correct 79% of the time. However, these predictions cannot be compared to the rankings directly, because the betting odds change dynamically during the course of the tournament, and therefore contain a lot of information that is not accessible prior to the start of the tournament. Table 2 provides an overview of the prediction accuracies of the different rankings and the betting odds, separated for each round of the tournament.

3.4. Individual recognition validities

Thus far we have analyzed rankings based on aggregated recognition, and we have shown that these rankings serve as useful predictors. However, in its original formulation, the recognition heuristic was introduced as a psychological model that describes a cognitive process (Goldstein & Gigerenzer, 2002), and as such, it applies to individual human beings making predictions based on their individual recognitions. Since the individuals in our study did not have access to aggregated recognition rankings, it remains unclear whether people can make accurate predictions based solely on their individual recognition knowledge.

To answer this question we first calculated the number of matches that each individual participant could have predicted using the recognition heuristic. The recognition heuristic can only be applied to pairs where the name of one player was recognized but not the other. If both or neither of the players are known, the recognition heuristic does not allow for a prediction. As the range of use depends on the percentage of recognized players, it differs among participants. For all of the decisions that each participant could make based on his or her recognition of the names, we then calculated the conditional probability of them making correct predictions, given that participants always used the recognition heuristic. This probability is commonly referred to as the recognition validity α.

On average, laypersons could have used the recognition heuristic to predict 21 of the 127 matches (17%), and out of these 21 games, using the recognition heuristic would have resulted in correct predictions for 15 matches ($\alpha=69\%$, SD=17%).[3] The average amateur could have used the heuristic to predict 50 matches (40%). Of these matches, 35 could have been predicted correctly based on the recognition heuristic ($\alpha=71\%$, SD=7%). Thus, even though laypeople

[2] The data stems from 5 international online betting sites, namely Bet365.com, Centrebet.com, Expekt.com, Interwetten.com, and Pinnacle Sports. We thank Joseph Buchdahl from tennis-data.co.uk for providing us with the data.

[3] The slight differences in the percentage figures from the absolute values are due to rounding.

Table 2
Percentage of correct predictions of the different rankings, and the betting odds for each round

	ATP		Recognition ranking			Wimbledon Seedings	Betting odds*
	Entry ranking	Champions race	Lay people	Amateurs	All participants		
1st round (64 matches)	.66	.65	.63	.69	.71	.63	.78–.80
2nd round (32 matches)	.69	.66	.63	.61	.59	.69	.72–.75
3rd round (16 matches)	.75	.94	.63	.63	.63	.78	.81
4th round (8 matches)	.75	.75	1.00	.88	1.00	.88	.75
Quarterfinals (4 matches)	.75	.75	1.00	.75	1.00	1.00	.75
Semifinals (2 matches)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Final (1 match)	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Total (all 127 matches)	.69	.70	.67	.68	.70	.70	.79

* As the number of correct predictions within the first two rounds differ slightly between the 5 online bookmakers, we report the range of correct predictions.

knew far fewer players than amateurs, they were about equally successful in predicting the outcomes of matches for which the heuristic was applicable.

To make a fair comparison between predictions made using the individual's recognition and its competitors ATP-CR, ATP-ER, and SEED, we calculated the percentage of correct predictions based on each of the competitors separately for those matches where the recognition heuristic could have been used by each participant. By doing this, we know, for each participant, whether he or she would have been better off deciding based on recognition, or based on the information from the competitors. For amateurs, the average percentage of correct predictions based on both ATP rankings is 70% (SD = 6%) in these cases, which is slightly lower than the recognition validity. SEED makes 71% (SD = 5%) and BET 79% (SD = 4%) correct predictions. Stated differently, from the 79 amateurs in our sample, 47 or 59% would have been equally well or better off deciding based on their recognition than based on ATP-CR. When compared with SEED, 40 amateurs (51%) would have been better off using their own recognition. Thus, for the matches for which the heuristic can be applied, most individual amateurs can be as accurate as, or even more accurate than, the Wimbledon experts and the predictions based on ATP rankings. For laypeople, individual recognition is not as accurate. Five laypeople could not use recognition at all, and for the remaining 100 laypeople, 80 would have been better off relying on ATP-CR, and 75 would have been better off relying on SEED. Table 3 gives an overview of the individual comparisons.

3.5. Accordance to the recognition heuristic

Given the relatively high recognition validity for both laypersons and amateurs, the question arises whether our participants actually made predictions in accordance with the heuristic.

For the 64 matches predicted in the first round, the average amateur could have used the recognition heuristic in 23.5 cases. Of these 23.5 cases, 20.9 decisions were in accordance with the prediction of the heuristic (89%). The average layperson could have used it in 10.3 cases, and 8.1 cases were in accordance (79%). For the 12 matches predicted in the 4th round and the quarterfinals, amateurs (on average) could have used the heuristic in 3.6 cases. Of these cases, 3.2 decisions (92%) turned out as the heuristic would have predicted. Laypersons could have used the heuristic in 2.7 cases,

Table 3
Percentage of correct predictions based on individual recognition and other predictors for all tennis matches for which the recognition heuristic was applicable

	Individual recognition validity α	Betting odds	ATP champions race	ATP entry ranking	Wimbledon seedings
Amateurs	.71 (SD = .07)	.79 (SD = .04)	.70 (SD = .06)	.70 (SD = .06)	.71 (SD = .06)
Laypeople	.69 (SD = .17)	.82 (SD = .11)	.78 (SD = .13)	.75 (SD = .14)	.75 (SD = .12)

and on average 2 decisions were in accordance (76%). Even though other researchers report even higher accordance rates (e.g., Serwe & Frings, 2006, report accordance rates of 88 to 93% and Ayton & Önkal, 2006, report accordance rates of 95%), in the present study concordances of amateurs and laypersons for both weeks were still quite high. Given that the recognition heuristic proved to be a valid predictor, this result is in line with the idea that our participants adaptively used a prediction strategy that worked well for the task they were facing. In the cases in which people decided against the recognition cue, and thus predicted that an unknown player would win over a known player, they were correct in 34% (average layperson) and 28% (average amateur) of the predicted matches respectively. As this is well below chance, whatever information this decision was based on had a lower validity than the recognition cue.

As has been pointed out by two anonymous reviewers, the fact that in many cases participants decided in accordance with the recognition heuristic is not direct evidence that participants actually used this heuristic. People might have come to the same decision by using a different decision strategy. Although the non-compensatory use of the recognition heuristic is debated, people at least incorporate the recognition cue into their judgments. For example, in Bröder and Eichler's (2006) study, a recognized city for which participants had further knowledge was chosen much more often when compared to an unrecognized city than when compared to another recognized city (without further knowledge). This reliable difference shows that recognition had an impact on their inferences.

3.6. Actual performance of laypersons and amateurs

It does not necessarily follow from a high validity that the participants in our study made accurate predictions across all matches. As mentioned above, the recognition heuristic can only be applied to cases in which one of the players is known and the other is not. If both players are known, participants might be able to use additional knowledge to make a prediction, but in cases in which neither of the players is recognized, participants can only make a guess. As a consequence, the actual accuracy may be very different from the recognition validity.

In the first round, amateurs on average recognized both players (and thus might have utilized additional knowledge) in 10 of the 64 matches. They did not recognize either of the players (and thus had to guess) for 29 matches. In the 4th round and the quarterfinals, amateurs on average recognized both players in 7 of the 12 matches and had to guess for 1 match. When the names of both players were known, amateurs made correct predictions in 75% (first round) and 79% (4th round and quarterfinals) of the matches. When the names of both players were unknown, amateurs were correct in 51% (first round) and 49% (4th round and quarterfinals) of the matches respectively.

Laypeople recognized far fewer players than amateurs, and thus had to guess more often. In the first round, laypeople on average only recognized both players in 1 of the 64 matches, and they had to guess the outcomes of 52 matches. In the 4th round and the quarterfinals, the average layperson recognized both players in 1 of the 12 matches, and had to guess in 8 cases. When the names of both players were known, the average layperson made correct predictions in 83% (first round) and 78% (4th round and quarterfinal) of the matches. When both player names were unknown, laypeople were correct in 49% (first round) and 60% (4th round and quarterfinals) of the matches. The 60% accuracy in the last case is significantly different from the 50% that would be expected for random guessing ($t[57]=3.6$; $p=.001$), which might be due to sloppy or incomplete completion of the recognition questionnaire in the field.

In summary, for the first 64 matches predicted in the first round, the average amateur predicted 61% correctly. For the 12 matches predicted in the 4th round and the quarterfinals, amateurs predicted 75% correctly. Laypeople predicted 52% (first round) and 61% (4th round and quarterfinals) of all matches correctly.

4. Discussion

Our results show that a ranking of tennis players based on aggregated name recognition by laypeople and amateurs was as effective in predicting match outcomes as official ATP rankings and Wimbledon experts' seedings. Also, for cases in which the recognition heuristic can be applied, individual decisions made based on mere name recognition are as accurate as predictions made by ATP rankings or Wimbledon experts. Our results are in line with those of Andersson et al. (2005), who showed that non-expert predictions of soccer games can be as successful as those of experts.

The idea that aggregating independent sources of information leads to better forecasts, even if the individual sources are not very accurate by themselves, has been raised by other researchers. Makridakis and Winkler (1983), as well as Winkler and Makridakis (1983), increased the accuracy of time series predictions by averaging across different forecasting models. In a study on medical decision making, Poses, Bekes, Winkler, Scott, and Copare (1990) found that when predicting patients' mortality risk, averaging across the opinions of rather inexperienced physicians (junior house officers in British hospitals) led to better estimates than the assessments of individual experienced physicians, even though the individual predictions of the inexperienced physicians were worse than those of the experienced ones. When forecasting the number of advertising pages sold by a news magazine, Hubbard Ashton and Ashton (1985) found that most of the improvement in accuracy is achieved by a simple average across three opinions. The influence of group size, individual accuracy, and the correlation between individuals on aggregated accuracy was explored in detail by Hogarth (1978). A comprehensive overview of the literature on combining forecasts was published by Clemen (1989).

Despite our use of larger samples, the inclusion of the names of all tennis players, and the prediction of more matches in two rounds, our results closely resemble the findings reported by Serwe and Frings (2006), who report an average accuracy of 72% for the aggregated amateur recognition and 66% for the aggregated recognition of laypeople in the 2003 tournament. For individual predictions, Serwe and Frings report a mean recognition validity of 73% for individual amateurs and 67% for individual laypeople respectively, which also matches the recognition validities in our study.

Although it is still possible that this match is a coincidence, the fact that two studies in different years and with different samples substantiate it, renders this unlikely. Hence, we conjecture that the relationship between recognition and success in sports might be more systematic than, for example, that between stock performance and recognition. But why might this be the case?

4.1. Ecological rationality of the recognition heuristic

In the present case, recognition is clearly not random, but rather is systematic, as it makes correct predictions in about 70% of the cases in which it can be applied. The fact that most decisions in our study are made in accordance with the recognition heuristic suggests that the participants implicitly understood that it is a useful strategy, and thus well adapted to the task they were facing. But why is there a relationship between name recognition and tennis performance at all? One explanation could be that the recognition validity is mediated by a third variable that relates to both individual recognition and the criterion to be predicted. If the criterion is a success in an international tennis tournament like Wimbledon, a potential mediator is mass media coverage. If the media report more on successful tennis players, there will be a correlation between the ability of a player and the number of times his name is mentioned. As this correlation between the criterion and a mediator is a property of the environment, it has been called an ecological correlation (Goldstein & Gigerenzer, 2002). If at the same time people recognize the players' names through the media, there will be a correlation between the mediator and recognition memory, the so-called surrogate correlation. To test whether media coverage really links the criterion to recognition, we counted how often the names of the 128 players were mentioned in both the sport section of a local newspaper (*Tagesspiegel*) and a national newspaper (*Süddeutsche Zeitung*) during the 12 months prior to the start of the competition. We then calculated the surrogate correlation between the newspaper coverage and the recognition ranking (RR-All), based on Goodman and Kruskal's γ-coefficients[4] (Gonzalez & Nelson, 1996). The surrogate correlation between RR-All and *Tagesspiegel* is $\gamma = .58$; between RR-All and *Süddeutsche Zeitung*, the γ-coefficient equals .59 (Fig. 1). We calculated the ecological correlation between the newspaper coverage and the success in the tournament based on a rule that was similar to the one used for the recognition rankings: predict that the player who is mentioned in the news more often will win the game, and do not count the cases in which the two players are mentioned equally often. The

[4] Because the calculation of Goodman and Kruskal's γ-coefficient is very similar to the calculation of the recognition validity α, one can easily be transformed into the other based on the following linear equation: $\gamma = 2\alpha - 1$.

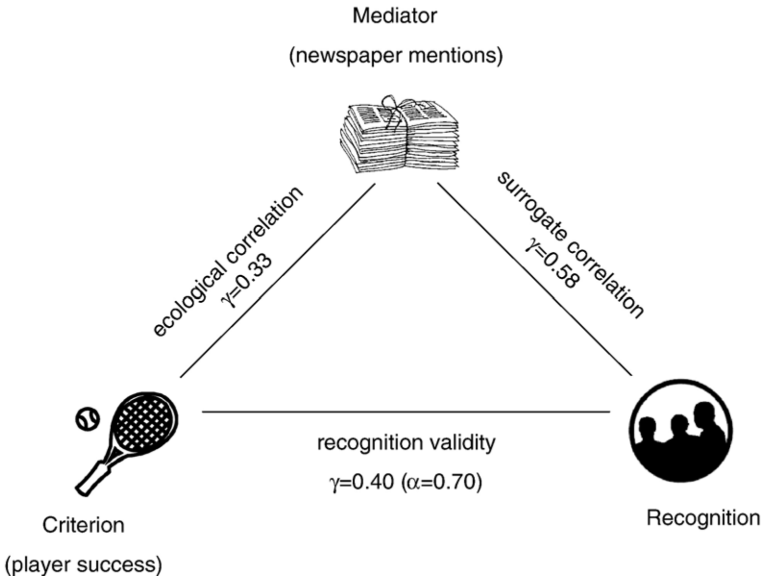

Fig. 1. The ecological rationality of the recognition heuristic.

resulting proportion of correct predictions can be linearly transformed into a γ-coefficient. For the two newspapers, the ecological correlation, as expressed in the γ-coefficients, equals .33 (α=.67). These results are not trivial, as players could be in the news, and thus recognized, for reasons unrelated to their ability on the tennis court, such as private matters or injuries.

In the Introduction we stated that, in order to make accurate predictions based on few pieces of information, a heuristic must be ecologically rational. That is to say, it needs to be suited to the structure of the environment in which it operates. In the domain of professional tennis, media coverage — and through this, also recognition — is correlated with success. This could explain why the recognition heuristic performs well, even though it makes predictions based on partial ignorance.

4.2. Betting odds

Among the competitors, the information contained in the betting market is the best predictor of success. This result is in line with Boulier and Stekler (2003), who found the same for the outcomes of NFL games. Similar results were also reported by Forrest, Goddard, and Simmons (2005), who showed that by the end of the soccer season, odds setters' forecasts were superior to tipsters', as well as to those of sophisticated statistical benchmark models. However, as mentioned above, betting odds change dynamically, and thus are not directly comparable to either the Wimbledon seedings or the recognition data.

5. Summary

Our results show that, when predicting the outcome of a sport event like the 2005 Wimbledon tennis tournament, relying on the mere recognition of player names by non-experts can be as accurate as forecasts based on expert ratings and on official ATP rankings. The systematic relationship between recognition and player success is presumably mediated by mass media coverage. Thus, by relying on their partial ignorance, non-experts are able to make accurate predictions by exploiting an environmental structure that contains relevant information and that is available at almost no cost.

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.

Ayton, P., & Önkal, D. (2006). *Effects of Ignorance and Information on Judgments and Decisions*. Unpublished manuscript, City University, London.

Borges, B., Goldstein, D., Ortmann, A., & Gigerenzer, G. (1999). Can ignorance beat the stock market? In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 59–72). New York: Oxford University Press.

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

Boulier, B. L., & Stekler, H. O. (2003). Predicting the outcomes of National Football League games. *International Journal of Forecasting*, *19*, 257−270.

Boyd, M. (2001). On ignorance, intuition, and investing: A bear market test of the recognition heuristic. *Journal of Psychology and Financial Markets*, *2*, 150−156.

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

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559−583.

Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In G. Gigerenzer, P. M. Todd, & The ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97–118). New York: Oxford University Press.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571−582.

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition*, *48*, 71−99.

Fildes, R., & Makridakis, S. (1995). The impact of empirical accuracy studies on time series analysis and forecasting. *International Statistical Review*, *63*, 289−308.

Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of the football betting market. *International Journal of Forecasting*, *21*(3), 552−564.

Forrest, D., & Simmons, R. (2000). Forecasting sport: The behaviour and performance of football tipsters. *International Journal of Forecasting*, *16*, 317−331.

Forster, M., & Sober, E. (1994). How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *British Journal for the Philosophy of Science*, *45*, 1−35.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*(3), 592−596.

Gigerenzer, G. (2004). Striking a blow for sanity in theories of rationality. In M. Augier & J. G. March (Eds.), *Models of a man: Essays in memory of Herbert A. Simon* (pp. 389−409).

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, *103*(4), 650−669.

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75−90.

Gonzalez, R., & Nelson, T. O. (1996). Measuring ordinal association in situations that contain tied scores. *Psychological Bulletin*, *119*, 159−165.

Hertwig, R., & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspectives on reasoning, judgment, and decision making* (pp. 213−231). Chichester, England: Wiley.

Herzog, S. (2005). *The boundedly rational fluency heuristic: How ecologically valid is recognition speed*? Unpublished Masters thesis, University of Basel, Switzerland.

Hogarth, R. M. (1978). A note on aggregating opinions. *Organizational Behavior and Human Performance*, *21*, 40−46.

Hubbard Ashton, A., & Ashton, R. H. (1985). Aggregating subjective forecasts: Some empirical results. *Management Science*, *31*(12), 1499−1508.

Kahnemann, D., & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer's critique. *Psychological Review*, *103*, 582−591.

Kahnemann, D., & Tversky, A. (2000). *Choices, values, and frames.* New York: Cambridge University Press.

Kareev, Y. (1995). Positive bias in the perception of covariation. *Psychological Review*, *102*, 490−502.

Kareev, Y. (2000). Seven (indeed, plus or minus two) and the detection of correlations. *Psychological Review*, *107*, 397−402.

Makridakis, S., Anderson, N. H., Carbone, R., Fildes, M., Hibon, R., Lewdowski, J., et al. (1982). The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting*, *1*, 111−153.

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., et al. (1993). The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting*, *9*, 5−22.

Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society*, *142*(2), 97−145.

Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, *16*, 451−476.

Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science*, *29*(9), 987−996.

Marewski, J. N., Gaissmaier, W., Dieckmann, A., Schooler, L. J., & Gigerenzer, G. (2005). Ignorance-based reasoning? *Applying the recognition heuristic to elections*. Paper presented at the 20th Biennial Conference on Subjective Probability, Utility and Decision Making, Stockholm.

Newell, B. R., & Fernandez, D. (2006). On the binary quality of recognition and the inconsequentiality of further knowledge: Two critical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, *19*, 333−346.

Newell, B. R., & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 923−935.

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

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

Poses, R. M., Bekes, C., Winkler, R. L., Scott, E., & Copare, F. J. (1990). Are two (inexperienced) heads better than one (experienced) head? Averaging house officers' prognostic judgments for critically ill patients. *Archives of Internal Medicine*, *150*, 1874−1878.

Richter, T., & Späth, T. (2006). Recognition is used as one cue among others in judgment and decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 150−162.

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, *135*, 207−236.

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

Shanteau, J. (1992). How much information does an expert use? Is it relevant? *Acta Psychologica*, *81*, 75−86.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology*, *41*, 1−19.

Snook, B., & Cullen, R. M. (2006). Recognizing national hockey league greatness with an ignorance-based heuristic. *Canadian Journal of Experimental Psychology*, *60*, 33−43.

Winkler, R. L., & Makridakis, S. (1983). The combination of forecasts. *Journal of the Royal Statistical Society, Series A*, *146* (2), 150−157.