

The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition

Stefan M. Herzog*

Ralph Hertwig†

Abstract

The collective recognition heuristic is a simple forecasting heuristic that bets on the fact that people's recognition knowledge of names is a proxy for their competitiveness: In sports, it predicts that the better-known team or player wins a game. We present two studies on the predictive power of recognition in forecasting soccer games (World Cup 2006 and UEFA Euro 2008) and analyze previously published results. The performance of the collective recognition heuristic is compared to two benchmarks: predictions based on official rankings and aggregated betting odds. Across three soccer and two tennis tournaments, the predictions based on recognition performed similar to those based on rankings; when compared with betting odds, the heuristic fared reasonably well. Forecasts based on rankings—but not on betting odds—were improved by incorporating collective recognition information. We discuss the use of recognition for forecasting in sports and conclude that aggregating across individual ignorance spawns collective wisdom.

Keywords: sports forecasting; rankings; betting odds; simple heuristics; name recognition; memory

“I do not believe in the collective wisdom of individual ignorance.” Thomas Carlyle¹ (1795–1881)

1 Introduction

With thousands of bookmakers accepting wagers on sporting events around the world, today, betting on sports is more popular than ever before. For example, in 2008 bettors in the UK alone wagered 980 million British pounds on soccer games—placing over 150 million bets in total (Gambling Commission, 2009). How should bettors and bookmakers make forecasts about sporting events? Many different approaches have been proposed (see e.g., Boulier & Stekler, 1999, 2003; Dixon & Pope, 2004; Goddard, 2005; Lebovic & Sigelman, 2001; Stefani, 1980). One common denominator is to muster plenty of knowledge—ranging from various indicators of the strength of individual players and teams to information about past outcomes, such as wins, losses—and then predict game scores (e.g., 3:2) or game outcomes (e.g., team A wins against team B; see e.g., Goddard & Asimakopulos, 2004) based on that knowledge.

We thank Thorsten Pachur, Benjamin Scheibehenne and Sascha Serwe for providing us with their raw data, Laura Wiles for editing the manuscript and the Swiss National Science Foundation for a grant to the first and second author (100014_129572/1).

*Corresponding author: Department of Psychology, University of Basel, Missionsstrasse 60-64, Basel, Switzerland. Email: stefan.herzog@unibas.ch.

†Department of Psychology, University of Basel

¹Cited in Menschel (2002), p. 136.

Knowledge about teams or players seems indispensable for rendering accurate forecasts—statistically or informally. Indeed, it seems absurd to assume that one can successfully predict which tennis player will win a match if one does not even know most of the names of his or her competitors in the tournament. Or can one? Surprisingly, there is mounting evidence that, contrary to Thomas Carlyle's intuition, the collective wisdom of individual ignorance genuinely exists. For instance, in a recent study, the ranks of tennis players performing in the Wimbledon 2005 tournament—based on how often they were recognized by 29 amateur tennis players—predicted the match winners better than the ATP Entry Ranking (Scheibehenne & Bröder, 2007; respondents recognized on average 39% of the players' names—thus respondents had far from complete knowledge). This “wisdom of ignorant crowds” is one among several examples in sports of the surprising predictive power of simple heuristics that forgo the exploitation of ample amounts of knowledge (Bennis & Pachur, 2006; Goldstein & Gigerenzer, 2009; Gröschner & Raab, 2006).

The fact that simple forecasting mechanisms can compete with or even outperform more sophisticated ones is by no means a new insight (e.g., Dawes, 1979; Makridakis & Hibon, 1979; see, e.g., Hogarth, in press, for a review). This finding, however, has been repeatedly met with resistance; is not widely put to use (see Armstrong, 2005; Goldstein & Gigerenzer, 2009; Hogarth, in press), and has not yet made it into popular textbooks of, for example, econometrics (see Hogarth, in press). One reason may be the intuitive appeal of the accuracy–effort

trade-off: The less information, computation, or time that one uses, the less accurate one's judgments will be. This trade-off is believed to be one of the few general laws of the human mind (see Gigerenzer, Hertwig, & Pachur, 2011), and violations of this law are seen as odd exceptions.

In the domain of forecasting sports events it is indeed difficult to judge to what simple forecasting strategies can outperform more complex ones simply because of the dearth of data. In a recent review, Goldstein and Gigerenzer (2009) noted that, "there is a need to test the relative performance of heuristics, experts, and complex forecasting methods more systematically over the years rather than in a few arbitrary championships" (p. 766). Focusing on the predictive power of collective recognition (or ignorance) in sports, this paper contributes to the literature in four ways. First, it presents two new studies on the predictive power of recognition in forecasting soccer games (World Cup 2006 and UEFA Euro 2008). These two studies will show to what extent the previous results can be replicated (see Evanschitzky & Armstrong, 2010; Hyndman, 2010, on the need of replicating findings in forecasting research). Second, it compares the predictive power of recognition in these two studies and in previously published research (reviewed in Goldstein and Gigerenzer, 2009) against two benchmarks in all tournaments: predictions based on official rankings (e.g., FIFA for soccer) and aggregated betting odds. Third, we investigate whether forecasts based on rankings and betting odds can be improved by incorporating collective recognition information. Fourth, we investigate the performance of a recognition-based heuristic that relies on the recognition of individual names rather than category names (e.g., the names of soccer players instead of the names of the soccer team itself).

Last but not least, let us emphasize that our investigation of collective recognition in the domain of sports should not be taken to mean that the power of collective recognition is restricted to this domain. Sports is just one illustrative domain; others are, for instance, prediction of political elections (e.g., Gaissmaier & Marewski, 2011), demographic and geographic variables (e.g., Goldstein & Gigerenzer, 2002).

2 The wisdom of ignorant crowds

Does more knowledge make for better forecasters? Research on the value of expertise in forecasting soccer games, for example, produced mixed findings: Some studies find that experts outperform novices (e.g., Pachur & Biele, 2007), some that they are equally accurate (e.g., Andersson, Edman, & Ekman, 2005; Andersson, Memmert, & Popowicz, 2009), and still others find that

novices can beat experts (e.g., Gröschner & Raab, 2006). Notwithstanding the question of when experts fare better relative to novices (see e.g., Camerer & Johnson, 1991), how is it possible that novices can ever outperform experts given that the former may not even recognize all the teams or players?

2.1 The benefits of ignorance

The key to this finding is that recognition or lack thereof is often not merely random, and thereby can reflect information valuable for forecasting. For example, successful tennis players are mentioned more often in the media than less successful ones, thus successful tennis players are more likely to be recognized by laypeople. As a consequence, the mere fact that a layperson recognizes one tennis player, but not another, carries information suggesting that the recognized one has been more successful in the recent past and thus is more likely to win the present game than the unrecognized one (Scheibehenne & Bröder, 2007).

More generally, whenever some target criterion of a reference class of objects (e.g., the size of cities, the salary of professional athletes, or the sales volume of companies) is correlated with the objects' exposure in the environment (e.g., high-earning athletes are more likely to be mentioned in newspapers; Hertwig, Herzog, Schooler, & Reimer, 2008), then the criterion will be mirrored in how often people recognize those objects (Goldstein & Gigerenzer, 2002; Pachur & Hertwig, 2006; Schooler & Hertwig, 2005). Consequently, recognition often allows reasonably accurate inferences in sports (for a review see Goldstein & Gigerenzer, 2009) and in many other domains (for a review see Pachur, Todd, Gigerenzer, Schooler, & Goldstein, in press).

Because experts recognize most—if not all—objects in their domain of expertise (almost by definition), they cannot fall back on partial ignorance as often as laypeople can (see Pachur & Biele, 2007, for an example in the soccer domain). Moreover, if the additional knowledge of experts fails to be more valid than the validity of mere recognition, then laypeople will be able to outperform experts in terms of accuracy (Goldstein & Gigerenzer, 2002; but see also Katsikopoulos, 2010; Pachur, 2010; Pleskac, 2007; Smithson, 2010).² But how can a forecaster benefit from the potential wisdom encapsulated in *collective ignorance*?

²How laypeople use recognition when making inferences is debated (see the view outlined in this and the previous special issue of *Judgment and Decision Making* on the recognition heuristic; for reviews of past research see Pachur, Bröder, & Marewski, 2008; Pachur et al., in press). This debate, however, does not pertain to our prescriptive analysis of recognition as a cue for forecasting heuristics.

2.2 Collective recognition heuristic: Using category versus individual names as input

A forecaster who wishes to predict—based on recognition—which of two contestants (e.g., tennis player, soccer team) will win a game can employ the *collective recognition heuristic* (adapted from Goldstein & Gigerenzer, 2009):

Ask a sample of semi-informed people to indicate whether they have heard of each contestant or not. Rank contestants according to their recognition rates (i.e., the proportion of people in the sample recognizing a contestant), and predict, for each game, that the contestant with the higher rank will win. If the ranks tie, guess.

The sample of people surveyed should be “semi-informed”; that is, they should recognize only a subset of the contestants, so that there is variability in the recognition rates, which—at least potentially—could predict the outcomes of interest. In contrast to semi-informed participants, experts are more likely to recognize all contestants, yielding many recognition rates of 100% and thus ranks that fail to differentiate between contestants.

It can, however, be hard to find semi-informed people for the following reason. With words that designate categories of things or beings, it can become difficult to discern those of which one has previously heard from those that one knows exist by logical deduction but has not heard of before. For example, has one heard before of the category of beings encompassing the *Bolivian* soccer team or does one “recognize” the category name based on the assumption that all South American countries have a national soccer team, and by extension, one must have heard of it? In contrast, it appears much easier to judge whether one has heard of a word that designates a particular thing (e.g., the Golden Gate Bridge) or a particular individual in the world (e.g., Roger Federer). A national soccer team can be seen as a *category* name, whereas its players can be seen as particular individuals within that category. If recognition of category words is more difficult and noisier than recognition of words designating particular individuals, then the performance of the collective recognition heuristic using the latter as input is likely to be better relative to the input in terms of category names. To investigate this possibility, we introduce the *atom recognition rate* that refers to the proportion of “atoms” (e.g., soccer players) recognized within a category (e.g., a soccer team). For instance, a person may recognize only one (4%) of the 23 players of the Bolivian team, relative to 10 (43%) players of the Brazilian team, but nevertheless (and correctly) judge that she has heard of both teams before.

Assessing the atom recognition rate instead of category recognition itself can be seen as a decomposition technique for recognition assessment (see MacGregor, 2001, on decomposition of quantitative estimates). Single-player sports are, by definition, “atomistic”. For example, tennis players are already atoms insofar as they cannot be decomposed into more meaningful, concrete subordinate components; here, category recognition and atom recognition overlap conceptually. In team sports, by contrast, players are the atoms from which their team is built. The collective recognition heuristic based on the atom recognition rate proceeds as follows:

Ask a sample of semi-informed people to indicate whether they have heard of each “atom” or not. Rank contestants according to their collective “atom” recognition rates (i.e., the mean atom recognition rate of each contestant across atoms and people surveyed), and predict, for each game, that the contestant with the higher rank will win. If the ranks tie, guess.

3 Method

3.1 Two performance benchmarks

3.1.1 Ranking rule

Rankings of players or teams based on their past performance are established and publicly accessible in many sports (e.g., FIFA ranking for soccer teams, ATP Entry Ranking for tennis players; Stefani, 1997). Higher-ranked players or teams—not surprisingly—tend to outperform lower-ranked ones (Boulier & Stekler, 1999; Caudill, 2003; del Corral & Prieto-Rodríguez, 2010; Klaassen & Magnus, 2003; Lebovic & Sigelman, 2001; Scheibehenne & Bröder, 2007; Serwe & Frings, 2006; Smith & Schwertman, 1999; Suzuki & Ohmori, 2008). In line with other researchers (e.g., Serwe & Frings, 2006; Suzuki & Ohmori, 2008), we use the accuracy of a *ranking rule* that predicts that the better-ranked team or player will win a game; if the ranks tie, the rule will guess. We use the most recent ranking published before the start of a tournament.

3.1.2 Odds rule

Betting odds are highly predictive of sport outcomes (e.g., Boulier, Stekler, & Amundson, 2006; Forrest & McHale, 2007; Gil & Levitt, 2007). We will use an *odds rule* that predicts that the team or player with the higher probability of victory (as revealed by aggregated odds) will win a game; if the odds tie, the rule will guess. We interpret the performance of this rule as an—admittedly

crude—approximation of the predictability of a tournament.³

There are three reasons why the odds rule will—in the long run—generally perform better than collective recognition and ranking rules, and thus represents an upper benchmark. First, betting markets are generally unbiased predictors of game outcomes (e.g., Sauer, 1998). Although bookmaker betting markets might not be completely efficient (e.g., Franck, Verbeek, & Nüesch, 2010; Vlastakis, Dotsis, & Markellos, 2009, for soccer bets), they are very effective in absorbing publicly available information (see Forrest, Goddard, & Simmons, 2005). Second, because bookmakers of online betting sites are allowed to update their odds right up until the start of each game, they can absorb very recent information. Betting odds thus have an informational advantage over strategies based on information that is “frozen” before the start of a tournament (Vlastakis et al., 2009)—such as recognition and rankings. Third, averaging odds over many different bookmakers has the advantage of canceling out strategic and unintentional inefficiencies of individual bookmakers (for a discussion about why different bookmakers’ odds may vary, see Vlastakis et al., 2009; for a discussion of the benefits of combining probability assessments, see e.g., Clemen & Winkler, 1999; Winkler, 1971; on the performance of aggregated odds to forecast soccer match results, see e.g., Hvattum & Arntzen, 2010; Leitner, Zeileis, & Hornik, 2010).

3.2 Comparing performance across studies

Different sports vary in terms of predictability. For example, outcomes of soccer and baseball games are less predictable based on a team’s past performance relative to ice hockey, basketball and American football (Ben-Naim, Vazquez, & Redner, 2006). Thus, the proportion of games predicted correctly can be directly compared across different strategies for a given tournament but not across different sports—or across different tournaments within the same sport, because even tournaments might differ in their predictability. To enable comparisons across different sports and tournaments, we introduce two performance measures that address those differences in predictability by taking into account the forecasts of a “gold standard” benchmark. We use aggregated betting odds as such a gold standard.

First, we analyze the *signal performance* of a strategy. This measure evaluates the proportion of correct forecasts of a strategy among those games where the gold standard

³We are aware of more sophisticated approaches to quantify parity and predictability of tournaments (e.g., Ben-Naim et al., 2006). Those measures, however, need to be calculated across large datasets of games and may not result in robust estimates for the considerably smaller sample sizes that we analyzed here.

(i.e., odds) predicted the winner of a game.⁴ The assumption is that the results of those games are less likely due to chance than those of games where the gold standard was wrong. The signal performance thus assesses a strategy’s ability to predict “what can be predicted” (i.e., true signals as opposed to noise). In doing so, this measure makes the performance of strategies across domains with different predictability (i.e., amount of noise) more comparable.

Second, we analyze the *normalized performance index (NPI)*. It expresses the performance of the target strategy as a fraction of the “gold standard” performance (i.e., odds) corrected for chance as follows:

$$NPI = \frac{\text{accuracy} - 50\%}{\text{gold standard performance} - 50\%}$$

We assume that the gold standard performance is larger than 50%, otherwise the NPI is either undefined (= 50%) or not interpretable (< 50%). An NPI of 0 indicates that the target strategy is at chance performance; a value of 1 indicates that it measures up to the gold standard. If a strategy scored, for example, 60% and the gold standard 70% correct predictions, the resulting NPI will be .5. Values above 1 indicate performance above the gold standard.

3.3 World Cup Soccer 2006 study

3.3.1 Participants

During the two days before the beginning of the tournament (8th and 9th June 2006), we obtained recognition judgments for each of the 23 players for all the 32 competing teams from 113 Swiss citizens approached on the University of Basel campus. Each participant judged a random third of all players. Participants’ age ranged from 20 to 53 years (*Mdn* = 24); 57% were female; 91% of participants were students.

3.3.2 Analysis

For each participant, the proportion of recognized players per team was calculated (atom recognition rate). Then for each team, the collective atom recognition rate was calculated by averaging participants’ values. We obtained the 2006 pre-tournament FIFA ranking⁵ of the teams (FIFA.com, 2010b) and aggregated 2006 pre-game betting odds (Betexplorer.com, 2010a). We then derived the predictions of the three strategies for the 48 group games.

⁴We thank an anonymous reviewer for this suggestion.

⁵Up to 2006, the FIFA ranking was based on the points received in international “A” matches during the last 8 years—giving more weight to more recent games. The points received for a match depended, among other things, on the importance of a match, the opponent’s strength, and the loss margin. After the World Cup Soccer 2006 the ranking system was changed and is now based only on the last 4 years—again giving more weight to more recent games (FIFA.com, 2010a).

3.4 UEFA 2008 study

3.4.1 Participants

During the five days before the beginning of the tournament (3rd to 7th June 2008), we obtained recognition judgments (for each of the 23 players for all the 16 competing teams, as well as for the 16 teams themselves) from participants recruited online (via email lists, online social networks, internet forums etc.). Of the 996 participants who started the study, 517 (52%) completed it and provided data amenable to analysis. Each participant judged a random third of all players and all 16 teams. Most participants were from Switzerland (39%) and Germany (19%); the remaining participants (42%) were from 38 different countries, each representing less than 10% of participants. Participants' age ranged from 12 to 74 years ($Mdn = 27$); 40% were female.

3.4.2 Analysis

For each participant the proportion of recognized players per team was calculated (atom recognition rate). Then for each team the collective atom recognition rate was calculated by averaging participants' values. We then assessed the collective recognition rate per team by calculating the proportion of participants recognizing a team. We conducted these calculations separately for the Swiss, German, and other-countries participants to explore regional differences in the performance of collective recognition and collective atom recognition⁶. We obtained the 2008 pre-tournament FIFA ranking of the teams (FIFA.com, 2010b) and aggregated 2008 pre-game betting odds (Betexplorer.com, 2010b). We then derived the predictions of the four strategies for the 24 group games.

3.5 General methodology

We analyzed the performance of the collective recognition heuristic and the benchmarks in our two studies and in three published studies on the predictive power of recognition in sports that Goldstein and Gigerenzer (2009) reviewed. Two of the latter studies investigated Wimbledon Gentlemen's Singles tennis tournaments: 2003 (Serwe & Frings, 2006) and 2005 (Scheibehenne & Bröder, 2007). Both studies used two rankings as benchmarks: the *ATP Champions Race Ranking* (based on the

⁶We published predictions of a variant of the collective atom recognition heuristic online (Archive.org, 2008). There, we pooled participants from all countries and excluded for each game participants belonging to either of the two countries competing. This procedure aimed at creating "agnostic" collective atom recognition rates that would be free from "home bias"; participants tend to be heavily exposed to the names of players from their country's teams and—of course—to the names of their country's team itself.

games from the current calendar year) and the *ATP Entry Ranking* (based on the games from the previous 52 weeks)⁷. Serwe and Frings (2006) used odds from a single bookmaker (expekt.com). Scheibehenne and Bröder (2007) used odds from five bookmakers (bet365.com, centrebet.com, expekt.com, interwetten.com, and pinna-clesports.com); we used the average of the five bookmakers.

One other study investigated the UEFA Euro 2004 soccer championship (Pachur & Biele, 2007). We collected 2004 pre-tournament FIFA rankings (FIFA.com, 2010a, 2010b) and aggregated 2004 pre-game betting odds (Betexplorer.com, 2010c). Using the studies' raw data and the data that we retrieved online, we calculated the performance statistics reported in Tables 1 and 2.

In the knock-out phase of a soccer tournament, the betting odds refer to the result at the end of regular time (90 minutes plus added time) and not to the final result of the game (possibly including extra time and penalty shooting). To ensure that the odds predict the actual winners of the games, we only included the group games in the soccer tournaments. In addition, we excluded soccer games that ended in a draw because the recognition-based heuristics and the ranking rule cannot predict a draw⁸.

4 Results and discussion

We first present the main results of our two new studies (Table 1) and then summarize the results across all studies (Tables 1 and 2).

4.1 The two new studies

4.1.1 World Cup Soccer 2006

The collective recognition heuristic based on atom recognition correctly predicted 31 (84%) of the 37 games—clearly outperforming the FIFA ranking (70%) and achieving three fourths of the odds rule's performance (95% correct; $NPI = 0.76$; Table 1).

⁷Both rankings are based on points awarded to the winner of a match; the number of points depends on the importance of the tournament, the stage in the tournament, and the ranking of the defeated player (Stefani, 1997). The two rankings differ in the window of matches that they consider. The Champions Race ranking is based on the games played in the current calendar year, whereas the Entry Ranking is based on games played in the last 52 weeks. Thus the Champions Race ranking is based on less and more recent information than the Entry Ranking—except at the end of a year when the two rankings coincide.

⁸If one were to include those games, then all strategies would fare worse because they cannot predict a draw. (The odds only predicted one drawn game among the 98 games analyzed. Because this game also ended in a draw, it was not included in our analyses.) However, this would not change the relative standing of the different strategies, which is the main focus of this investigation. Generalizing the strategies so that they can predict draws (e.g., by introducing a just-noticeable difference between the two predictor values) is beyond the scope of this paper.

Table 1: Soccer tournaments: Performance of different forecasting strategies.

Tournament	Games (draws)	Population	N	Performance				Applicability	
				Collective recognition heuristic	Collective atom recognition heuristic ^a	FIFA ranking rule	Odds rule	Collective recognition heuristic	Collective atom recognition heuristic ^a
				UEFA Euro 2008	24(3)	Swiss German International	202 99 216	65% ^{60%} _{0.71} 81% ^{60%} _{0.71} 85% ^{69%} _{1.36}	77% ^{62%} _{0.83} 85% ^{62%} _{0.83} 85% ^{62%} _{0.83}
UEFA Euro 2004	24(8)	Berlin	121	88% ^{66%} _{0.63}	—	92% ^{75%} _{1.00}	75%	94%	—
World Cup 2006	48(11)	Basel	113	—	86% ^{84%} _{0.76}	74% ^{70%} _{0.45}	95%	—	100%

Note. *N* denotes number of participants. The percentages indicate the proportion of non-drawn games predicted correctly by a strategy (“Performance”) and the proportion of non-drawn games where the recognition-based heuristics were applicable (“Applicability”). The superscripts indicate the proportion of non-drawn games predicted correctly by a strategy only for those games that were correctly predicted by the odds rule (signal performance). The subscripts indicate the normalized performance index (NPI; see Method section for details).

^aEach participant indicated recognition judgments for a random third of the 23 players’ names.

4.1.2 UEFA Euro 2008

The collective recognition heuristic based on the Swiss, German, and other participants’ recognition of team names (or lack thereof) predicted 12.5 (60%), 12.5 (60%), and 14.5 (69%) of the 21 games correctly⁹—outperforming the FIFA ranking (57%) and achieving between 0.71 and 1.36 of the odds rule’s performance (64% correct). The collective recognition heuristic based on recognition of the players’ names (atom recognition) correctly predicted 13 (62%) of the games for all three subsets of participants—outperforming the FIFA ranking (57%) and achieving 0.86 of the odds rule’s performance. In this tournament, the collective recognition heuristic based on recognition of individual names did not fare better than the recognition heuristic based on team names (see Table 1).

4.2 Results across all studies

The names of tennis players already designate individuals rather than categories, therefore the distinction between category recognition and atom recognition disappears in the domain of tennis. Table 2 reports the performance statistics for the two tennis tournaments across strategies. Across soccer and tennis tournaments (Tables 1 and 2), the collective recognition heuristic based on the names of individual soccer or tennis players outperformed the ranking rules in six comparisons, tied in

one and yielded in five comparisons. The signal performance of the collective recognition heuristic ranged from 66% to 86% (*Mdn* = 78%, *CI*¹⁰ [.73, .85])—that of the ranking rules from 69% to 92% (*Mdn* = 75%, *CI* [.72, .85]). Not surprisingly, the odds rule outperformed the collective recognition heuristic in all eight comparisons; it also beat the ranking rules in six out of seven comparison and tied in the remaining one. The collective recognition heuristic’s normalized performance indices (NPIs) in the eight tournaments ranged from 0.49 to 0.83 (*Mdn* = 0.76, *CI* [0.58, 0.83])—that is, the collective recognition heuristic achieved, on average, about three fourths of the odds rules’s performance. As a comparison, the NPIs of the ranking rules ranged from 0.45 to 1.00 (*Mdn* = 0.62, *CI* [0.49, 0.79]).

The collective recognition heuristic based on team names (in the soccer tournaments, see Table 1) outperformed the ranking rule in three of four comparisons and yielded signal performance measures of 65%, 81%, 85%, and 88%. In three out of four cases, the odds rule performed better than the collective recognition heuristic (NPIs: 0.63, 0.71, 0.71 and 1.36).

Comparing the variability in performance of all strategies in the soccer (Table 1) and the tennis tournaments (Table 2) reveals that the results in tennis seem to be more stable than those in soccer. One possible reason is that the latent “real” competitiveness of tennis players is more reliably assessed than that of soccer teams for two reasons. First, the tennis tournaments feature a larger set of games

⁹Whenever a strategy was tied on its predictors, we counted that game as 0.5 correctly predicted.

¹⁰The 95% confidence interval of the median was calculated using Wilcox’s (n.d., 2005) function *sint*.

Table 2: Tennis tournaments: Performance of different forecasting strategies.

Tournament	Matches	Population	N	Performance				Applicability
				Collective recognition heuristic	ATP Race Ranking rule	ATP Champions Ranking rule	ATP Entry Ranking rule	Collective recognition heuristic
Wimbledon 2005	127	Amateur players (Berlin)	79	^{78%} 68% _{0.63}	^{82%} 70% _{0.71}	^{81%} 69% _{0.67}	79%	99%
		Laypeople (Berlin)	105	^{74%} 67% _{0.59}				98%
Wimbledon 2003	96	Amateur players (Duisburg)	29	^{77%} 72% _{0.76}	^{75%} 68% _{0.62}	^{73%} 66% _{0.55}	79%	100%
		University students (Jena)	96	^{66%} 64% _{0.49}				94%

Note. *N* denotes number of participants. The percentages indicate the proportion of games predicted correctly by a strategy (“Performance”) and the proportion of games where the recognition-based heuristics were applicable (“Applicability”). The superscripts indicate the proportion of games predicted correctly by a strategy only for those games that were correctly predicted by the odds rule (signal performance). The subscripts indicate the normalized performance index (NPI; see Method section for details).

than the soccer tournaments and, second, within a tennis match there are more opportunities for the latent skill to reveal itself than in a soccer game (i.e., many more serves and points in tennis than goal opportunities and actual goals in soccer).

To put the performance of recognition into perspective, it is illustrative to compare it to the performance of the recognition heuristic in domains outside sport. The proportion of correct forecasts based on collective (atom) recognition ranged between 60% and 84% across the 12 samples analyzed in this paper (*Mdn* = 65%, *CI* [.62, .69]). Similarly, people’s median individual recognition validity (i.e., the median proportion of times the recognition cue made a correct prediction based on an individual’s recognition knowledge among all non-drawn games) ranged between 56% and 79% (*Mdn* = 67%, *CI* [.59, .71]; see Tables 3 and 4). In five representative environments investigated by Hertwig et al. (2008), the recognition validities ranged from 61% (cumulative record sales of music artists), 67% (wealth of billionaires), 69% (earnings of athletes), 70% (revenue of German companies) to 83% (population size of U.S. cities). This comparison suggests that the predictiveness of recognition may be comparable in the domains of sport, economics, and geography.

4.3 The benefits of aggregating ignorance

The collective recognition and the collective atom recognition heuristic use the aggregated ignorance of a group of people to make predictions. In contrast, the recognition heuristic uses the recognition knowledge of a single person (Goldstein & Gigerenzer, 2002). But why aggregate? The benefits of aggregating ignorance are two-fold.

First, it increases the applicability of recognition-based heuristics (that is, the proportion of cases where a prediction can be made) and thus reduces the proportion of cases where the heuristic resorts to guessing because both objects have the same recognition value. Tables 3 and 4 summarize several measures calculated on the level of individual participants for the soccer and tennis tournaments: the recognition rate (i.e., proportion of team or player names recognized), the applicability rate (i.e., proportion of games where the recognition cue was not tied; that is, where it allowed a prediction), the recognition accuracy (i.e., the proportion of correct forecasts, assuming that a forecaster guesses when the recognition cue is tied), and the recognition validity (i.e., the proportion of correct forecasts only for those games where the recognition cue was *not* tied; see Goldstein & Gigerenzer, 2002). As can be seen in Tables 1 to 4, in all 12 samples in this study, the applicability of the collective heuristics was higher than that of the participants’ individual heuristic (i.e., applicability of the recognition heuristic). This difference is most pronounced for the collective recognition heuristic in the UEFA Euro 2008 tournament. Here, the median participant recognized all names of the soccer teams (see Table 3) and thus could never apply the recognition heuristic, whereas the collective recognition heuristic could be applied in almost all games (see Table 1). In contrast, because an individual’s *atom* recognition rate for a soccer team can take graded values between 0 and 1, the individual atom recognition heuristic could be applied almost as often as the collective atom recognition heuristic (86% for the median participant vs. 100% for the collective atom recognition heuristic, see Tables 1 and 3).

Table 3: Soccer tournaments: Measures for individual participants.

Tournament	Population	N	Target name	Recognition rate		Applicability rate		Recognition accuracy		Recognition validity	
				Mdn	[95% CI]	Mdn	[95% CI]	Mdn	[95% CI]	Mdn	[95% CI]
UEFA Euro 2008	Swiss	202	Team	100%	[97, 100]	0%	[0, 10]	50%	[50, 50]	63%	[57, 67]
			Player ^a	24%	[22, 28]	86%	[86, 90]	55%	[55, 60]	58%	[56, 61]
	German	99	Team	100%	[94, 100]	0%	[0, 11]	50%	[50, 50]	59%	[50, 67]
			Player ^a	24%	[19, 31]	86%	[81, 90]	55%	[52, 57]	56%	[53, 61]
	International	216	Team	100%	[100, 100]	0%	[0, 0]	50%	[50, 50]	67%	[60, 77]
			Player ^a	27%	[24, 31]	86%	[86, 90]	57%	[55, 60]	60%	[60, 63]
UEFA Euro 2004	Berlin	121	Team	69%	[63, 75]	48%	[38, 50]	56%	[56, 59]	71%	[70, 80]
World Cup 2006	Basel	113	Player ^a	11%	[9, 14]	70%	[59, 78]	69%	[66, 70]	79%	[77, 82]

Note. *N* denotes number of participants. Measures reported in this table: recognition rate (i.e., proportion of names recognized), applicability rate (i.e., proportion of games where the recognition cue was not tied; that is, where it allowed a prediction), recognition accuracy (i.e., the proportion of correct forecasts, assuming that a forecaster guesses when the recognition cue was tied) and recognition validity (i.e., the proportion of correct forecasts only for those games where the recognition cue was not tied). All calculations are only based on the non-drawn games. The group distributions are summarized by the median because many of them were highly skewed. The 95% confidence intervals of the median are calculated using Wilcox’s (n.d., 2005) function *sint*.

^aEach participant indicated recognition judgments for a random third of the 23 players’ names.

The second benefit of aggregating recognition judgments is that it creates a “portfolio of ignorance”. People may recognize a team or a player for reasons that are unrelated to the team’s or player’s competitiveness (e.g., a widely discussed extramarital affair; or because the name is a common name, or because of random error in the recognition judgment; see also Pleskac, 2007). To the extent that different people’s recognition knowledge represents different “errors”, those errors will tend to cancel out when aggregating recognition judgments; this benefit of error cancellation by aggregation has been widely discussed in the forecasting (e.g., Armstrong, 2001; Clemen, 1989) and machine learning literature (e.g., Dietterich, 2000). As an illustration of the benefit of error cancellation, consider recognition of the names of soccer players in the UEFA Euro 2008 tournament. We compared the accuracy of an individual participant’s recognition heuristic (i.e., recognition validity) with the accuracy of the collective atom recognition heuristic for only those games where this participant’s recognition knowledge allowed a prediction. The recognition validity of the majority of Swiss (72%, CI¹¹ [.65, .78]), German (79%, CI [.70, .86]) and international participants (72%, CI [.65, .77]) was lower than the accuracy of their individually matched collective atom recognition heuristic. This superiority of

collective atom recognition reflects error cancellation and not a higher applicability of the collective heuristic.

4.4 Does collective recognition improve the forecasts based on rankings and betting odds?

The collective recognition heuristic enables predictions that are on par with those of official rankings in the studies analyzed. One could therefore conclude that rankings should be preferred to collective recognition because the former are easier to obtain than the latter (see the general discussion for a broader discussion of this topic). But could it be that collective recognition contains predictive information that goes beyond that contained in rankings? That is, could one combine rankings with collective recognition and arrive at predictions that are superior to those based on rankings alone? Furthermore, could collective recognition similarly improve forecasts based on betting odds?

To answer these questions, we compared regression models of the strategies proper (i.e., collective recognition heuristic, ranking rule, and odds rule), relative to regression models combining recognition with rankings and odds, respectively. Specifically, we conducted a series of logistic (logit) regression models that was built on the following logic (see del Corral & Prieto-Rodríguez, 2010): For each of the strategies proper, we defined a

¹¹The 95% confidence interval of a binomial proportion was calculated using Wilcox’s (n.d.) function *acbinomci* (see Brown, Cai, & Das-Gupta, 2002).

Table 4: Tennis tournaments: Measures for individual participants.

Tournament	Population	N	Recognition rate		Applicability rate		Recognition accuracy		Recognition validity	
			Mdn	[95% CI]	Mdn	[95% CI]	Mdn	[95% CI]	Mdn	[95% CI]
Wimbledon 2005	Amateur players (Berlin)	79	50%	[38, 54]	40%	[35, 42]	59%	[57, 60]	73%	[70, 78]
	Laypeople (Berlin)	105	8%	[7, 9]	11%	[8, 14]	51%	[51, 52]	70%	[67, 75]
Wimbledon 2003	Amateur players (Duisburg)	29	37%	[31, 47]	41%	[38, 45]	58%	[58, 59]	70%	[69, 73]
	University students (Jena)	96	5%	[4, 8]	13%	[9, 17]	52%	[51, 54]	67%	[65, 72]

Note. *N* denotes number of participants. Measures reported in this table: recognition rate (i.e., proportion of names recognized), applicability rate (i.e., proportion of games where the recognition cue was not tied; that is, where it allowed a prediction), recognition accuracy (i.e., the proportion of correct forecasts, assuming that a forecaster guesses when the recognition cue was tied) and recognition validity (i.e., the proportion of correct forecasts only for those games where the recognition cue was not tied). The group distributions are summarized by the median because many of them were highly skewed. The 95% confidence intervals of the median are calculated using Wilcoxon's (n.d., 2005) function *sint*.

measure (explained below) indicating how strongly the strategy favored what it determined to be the winner. Using these measures, we next determined whether the strategies were indeed more likely to be right when they had a stronger favorite. Reiterating the same procedure, we finally analyzed whether the performance of the ranking and the odds rule improved when recognition was added as an additional predictor. Because of the small number of games in the soccer tournaments and the heterogeneity of the strategies' performance (see Table 1), making it impossible to pool across tournaments, we did not obtain robust results for this domain. The following analysis thus only concerns the tennis tournaments. To simplify the analyses, we averaged the two ATP rankings (Champions Race Ranking and Entry Ranking) into one overall ATP ranking and pooled the two tournaments (including a dummy variable coding for the games of the 2005 tournament) in all regression models. We also averaged the collective recognition rates from the experts and laypeople before computing the collective recognition rankings. Separate analyses for the two tournaments, the two rankings, and the two participant pools (experts vs. laypeople) yielded qualitatively similar results.

In the analyses, we used the log ratio of the ATP rankings—lower-ranked player divided by the higher-ranked player—as a measure of how strongly the ranking rule predicted the win to occur. This log ratio successfully predicts the probability that a better-ranked tennis player defeats a lower-ranked player (see e.g., del Corral & Prieto-Rodríguez, 2010, for an analysis of 4,064 Grand Slam tennis matches from 2005 to 2008). For collective recognition, we ranked the players according to their collective recognition rates and also used the log ratio of the ranks: lower-ranked player divided by the higher-

ranked player. Those two log ratio measures imply that the same absolute difference in ranks is—by taking the ratio—more important the higher ranked both players are and that the importance of the proportional difference between two ranks is subject to—by taking the logarithm—diminishing marginal increases.

Betting odds can be understood as revealed probability judgments and can be converted into “as-if” probabilities by taking the reciprocal of the decimal odds (see e.g., Vlastakis et al., 2009, eq. 2). We calculated these probabilities, made sure that they add up to 1 for each game—their sum is smaller than 1 because bookmakers want to ensure a stable income from the margin (Vlastakis et al., 2009)—and then calculated odds ratios conditioned on the player with the better odds of winning the game. Because the odds ratios were strongly skewed, we used log odds ratios for the analyses.

We ran a baseline model for each of the three strategies that predicted whether or not the strategy's forecast was correct based on the respective strategy's predictor variable (“ATP.win ~ ATP”, “Odds.win ~ Odds” and “REC.win ~ REC”). Two models (“ATP.win ~ ATP + REC” and “Odds.win ~ Odds + REC”) tested to what extent the addition of collective recognition rankings improved accuracy, relative to the ATP ranking and the odds alone. For the latter two models, the ratio of the recognition rankings needs to be defined in the same way as the respective target ratio (ATP and Odds): That is, we divided the recognition ranking of the player with the worse ATP ranking (worse odds) by the recognition ranking of the player with the better ATP ranking (better odds).

Table 5 reports model coefficients, the Bayesian Information Criterion (BIC; Raftery, 1995) and Brier scores

Table 5: Tennis tournaments: Analysis of the additional predictive utility of collective recognition.

Model	BIC	Coefficients					Brier score		
		Intercept	ATP	Odds	REC	2005	All	Fit	Test
ATP.win ~ ATP	281.7	0.10 [-0.49, +0.69]	0.50 [0.19, 0.85]	—	—	0.14 [-0.45, +0.73]	.205	.203	.212
ATP.win ~ ATP + REC	277.6	0.19 [-0.41, +0.79]	0.26 [-0.10, +0.64]	—	0.48 [0.17, 0.80]	0.15 [-0.45, +0.75]	.194	.192	.204
Odds.win ~ Odds	222.9	0.26 [-0.41, +0.94]	—	0.73 [0.41, 1.10]	—	-0.22 [-0.92, +0.46]	.151	.152	.158
Odds.win ~ Odds + REC	228.3	0.26 [-0.41, +0.94]	—	0.73 [0.37, 1.13]	0.01 [-0.35, +0.36]	-0.22 [-0.92, +0.46]	.151	.151	.161
REC.win ~ REC	282.2	0.55 [0.02, 1.09]	—	—	0.43 [0.05, 0.87]	-0.16 [-0.76, +0.43]	.203	.202	.211

Note. Logistic regression analyses predicted whether a strategy correctly forecast the winner of a game (ATP.win, Odds.win and REC.win) based on a subset of the following predictors (see main text for details): log ratio of ATP rankings (ATP), log odds ratio (Odds), log ratio of recognition rankings (REC), and a dummy variable coding for the games of the Wimbledon 2005 tournament. The reported coefficients are unstandardized; 95% confidence intervals are reported in square brackets. Brier scores are reported for the full dataset (“All”), as well as for the learning dataset (“Fit”) and the test dataset (“Test”) in the cross-validation simulation (100,000 samples; see main text for details). The standard errors of the Brier scores in the cross-validation simulation were smaller than .00011. Random probability forecasts drawn from a uniform distribution ([0, 1]) yielded a Brier score of .332; lower Brier scores imply better probability forecasts.

(Brier, 1950; Yates, 1982, 1994)—a measure of the quality of probabilistic forecasts where lower values indicate better forecasts.¹² We ran a cross-validation simulation where we fitted the five models to a random two thirds of the games and then—using the fitted parameters—predicted the outcomes of the remaining third; we repeated that procedure for 100,000 cross-validation samples. Table 5 reports three Brier scores for each model: the score based on the full sample (column “All”) and the average scores for the learning dataset (column “Fit”) and the test dataset (column “Test”) across all cross-validation samples. The standard errors of the Brier scores in the cross-validation simulation were smaller than .00011.

Four results emerged. First, the larger the differences between the ranks or odds of two players, the more likely that the strategy’s forecast was correct, as indicated by the positive slopes of the predictors in the three baseline models. The slopes in a logit regression model can be converted into odds ratios of a “unit change” on the predictor variable by plugging the slopes into the exponen-

tial function. For the ATP model, for example, the odds of the better-ranked player winning against the lower-ranked player are $e^{0.50}$; that is, 1.66 times higher for a pair of players with a log ratio that is one unit larger than that of another pair of players. The respective odds ratios are 2.08 and 1.54 for the log odds ratios of the betting odds and the log ratios of the collective recognition rankings, respectively.

Second, whereas the probability forecasts of the ATP rankings and the collective recognition rankings were comparable in terms of the cross-validated Brier scores (.212 and .211), those of the betting odds were clearly superior (.158). The recognition model yielded a better Brier score, relative to the ATP model’s Brier score, in only 52% of the cross-validation samples. In contrast, the odds model yielded a better score, as compared with both the ATP and the recognition model, in 99% of the samples. The BIC of the odds model is 59 units lower than that of the other two models, which indicates “very strong” evidence in support of the odds model (see Raftery, 1995, pp. 138–139).

Third, adding recognition rankings to the ATP rankings improved forecasts relative to the ATP rankings only: the cross-validated Brier score dropped from .212 to .204. The combined model achieved a better score in 82%

¹²The Brier score is defined as the average squared difference between the predicted probability that an outcome occurs and an indicator variable; the latter is 1 if the event occurs, and 0 otherwise. The score ranges between 0 and 1; smaller values indicate better forecasts.

of the cross-validation samples. The BIC decreased by 4.0—indicating that the data are roughly 8 times ($e^{4.0/2} = 7.56$) more likely assuming the combined model as compared to the ATP model. Assuming that both models are equally likely a priori, this implies a posterior probability of the combined model of 88% (see Wagenmakers, 2007, pp. 796–797).

Fourth, adding recognition rankings to the betting odds did not improve forecasts relative to odds only. It actually led to worse forecasts. The cross-validated Brier score increased from .158 to .161. The combined model achieved a worse score in 62% of the cross-validation samples. The BIC increased by 5.4, indicating that the data are roughly 15 times ($e^{5.4/2} = 14.92$) more likely assuming the simple as compared to the combined model. The posterior probability of the simple model is 94%, assuming equal priors.

5 General discussion

Our replications and analyses of previous studies have yielded four major findings. First, in the three soccer and the two tennis tournaments the collective recognition heuristic enables forecasts that consistently perform above chance, and that are as accurate as predictions based on official rankings (Tables 1 and 2). Second, we compared the performance of the collective recognition heuristic based on the recognition of category names (the soccer team's name) and names of individual soccer players for the UEFA Euro 2008 tournament and did not find appreciable differences in their performance (Table 1). Apparently in this tournament, the recognition of category words is no less reliable or valid than the recognition of words designating particular individuals. Third, aggregated betting odds, on average, are superior to predictions based on rankings or collective recognition (Tables 1, 2, and 5). This result, however, was to be expected due to the informational advantage of betting odds (see e.g., Vlastakis et al., 2009). Fourth, in the two tennis tournaments, the collective recognition heuristic, the ATP and the odds rule were more likely to render correct forecasts the larger the differences on their respective predictors. This implies that the larger the difference in the ranks of, for example, recognition rates, the more confident a forecaster can be in her predictions. Moreover, the forecasts of the ATP rule—but not those of the odds rule—can be improved by incorporating collective recognition rankings into the forecast.

5.1 When should one use the wisdom of ignorant crowds?

In domains where established and valid rankings or betting odds are available, the most straightforward approach seems to use those rankings or odds to render forecasts. The effort of collecting recognition judgments does not seem to pay off when those alternative—already conveniently pre-calculated—cues are available. In practice, however, the collective (atom) recognition is still an attractive option for at least three reasons.

First, in some domains forecasters might not trust the predictive ability of a ranking system because they may feel that the logic behind the system is partially flawed. For example, up to the World Cup 2006, the FIFA ranking was based on games from the last 8 years and many commentators felt that it did not adequately reflect the *current* strength of the teams (BBC Sport, 2000). The ranking system was later revised to only encompass the last 4 years (FIFA.com, 2010a). In addition, some ranking systems—by their very design—may reflect more than merely the latent skills of the contestants. For example, because the ATP ranking system awards more points for matches in more prestigious tournaments (Stefani, 1997), there is an incentive to play many matches in such tournaments. These and other incentives may lower a ranking's ability to predict future winners. Second, as our analysis of the two tennis tournaments suggests, the predictions based on ranking information may be improved by incorporating collective recognition information. Such a combined use of rankings and collective recognition is especially attractive when forecasters are unsure about the trustworthiness of the ranking system and would like to diversify the risk of relying on bad information by including additional, non-redundant information into their predictions (see also Graefe & Armstrong, 2009, on a combined use of recognition-like information, rankings, and betting odds in tennis tournaments). Third, betting odds might not be available at the time when forecasters render their predictions. In sports, betting odds are usually only available for those games for which it is known who will play whom. At the start of tournaments with a later knock-out phase (e.g., UEFA Euro and World Cup Soccer tournaments), one can only bet on the outcomes of the round-robin games, but not on the later knock-out phase because it is not yet known who will encounter whom. Only when the tournament moves to the next stage will bookmakers offer new bets on those games.

The results of our analyses suggest that in the domains of soccer and tennis—and possibly also in other domains—collective (atom) recognition can be expected to achieve about three fourths of the performance of ag-

gregated betting odds and to be on par with official ranking systems. Thus when rankings and odds are not trustworthy or available, collective recognition is an alternative and frugal forecasting option.

But when should one *not* use collective recognition and switch to other approaches? People's recognition knowledge mirrors how often they encountered names (e.g., Goldstein & Gigerenzer, 2002; Hertwig et al., 2008) and the probability of encountering a particular name partly depends on how "important" that name is in people's environment (e.g., people write and read, on average, more about successful companies and athletes than about less successful ones; Hertwig et al., 2008; Scheibehenne & Bröder, 2007). We can thus expect recognition generally to be a valid cue in the domain of sports and in many other domains in which the criterion dimension (e.g., size, wealth, or success) matters to the public. By the same token, however, one should refrain from using collective recognition for obscure criteria that are of little interest to people and where there thus will be no correlation between the criterion and recognition (e.g., shoe size of tennis players and their name recognition; see also Pohl, 2006).

5.2 Whom to ask and how many?

If a forecaster decides to use the collective (atom) recognition heuristic, two main questions arise: Whom to ask and how many? Regarding the first question, forecasters should collect responses from a diverse set of respondents that have been exposed to different information environments. In the same way that, for example, economic experts from different schools of thought (and thus likely exposed to different information and assumptions) have errors that are less correlated than those of experts from the same school of thought (Batchelor & Dua, 1995), the errors in recognition judgments from a diverse set of people may also be less correlated than the errors of similar people. This means that errors are more likely to cancel out with a diverse set of people. The finding that the collective recognition heuristic fared better with recognition judgments stemming from respondents from all over the world than with recognition judgments stemming from Swiss or German respondents in the UEFA Euro 2008 tournament highlights the importance of non-redundant recognition judgments. The prescription of using recognition data from different sources mirrors Armstrong's (2001) principle of using "different data or different methods" (p. 419) when combining forecasts.

How many people should you survey? This question can be rephrased as: How large should the sample size be so that the estimates of the true recognition rates are

reasonably reliable? Because the benefit of adding an additional binary observation (i.e., recognized the name vs. did not recognize the name) in terms of accurately assessing the population value decreases with increasing sample size, we suspect that most of the gains in predictive power can be achieved with a few dozen observations. When using atom recognition, the necessary sample size might be even lower because estimation error will already cancel out when aggregating the atom recognition rates within a category (e.g., from the player names to the soccer team).

5.3 How can one use the wisdom of ignorant crowds even when there is no crowd available?

Given the predictive advantage of aggregating ignorance, how could a single forecaster still profit from a crowd's ignorance even when no crowd is available? We recently showed that individual people can simulate a "crowd within" to improve their quantitative judgments using *dialectical bootstrapping* (Herzog & Hertwig, 2009)—thus emulating a social heuristic (see Hertwig & Herzog, 2009): Canceling out error by averaging their first estimate with a second, dialectical one that uses different assumptions and is thus likely to have an error of different sign. We speculate that individual forecasters could simulate the "wisdom of ignorant crowds" within their own mind by, for example, estimating the proportion of people among a specified reference class (e.g., one's family and friends or a representative sample of residents from a country) who would recognize team or player names. In the same way, however, that the errors of two different people's estimates are more independent than the errors of two estimates from the same person (e.g., Herzog & Hertwig, 2009), we suspect that recognition knowledge from different people is more independent than the recognition knowledge of a simulated crowd.

Another approach is to look for proxies of people's recognition knowledge. Frequencies of name mentions in large text corpi (e.g., number of hits on google.com or in online newspaper archives) are good proxies of recognition data (see e.g., Goldstein & Gigerenzer, 2002; Hertwig et al., 2008) and very easy and quick to collect. Predicting for the Wimbledon 2005 tournament, for example, that a game will be won by the tennis player mentioned more often in the sports section of the German newspapers *Tagesspiegel* or *Süddeutsche Zeitung* (during the 12 months prior to the start of the tournament) was almost, but not quite as predictive as collective recognition (Scheibehenne & Bröder, 2007). Also,

the frequency with which users enter names into search engines—another proxy for how well known and important objects are—can be used to predict events. For example, across the 1,016 matches of the eight Grand Slam tennis tournaments in 2007 and 2008, the tennis player who was searched for more often won 70% of the games (Graefe & Armstrong, 2009). As a comparison, a ranking rule (based on the ATP Entry Ranking) predicted 72% and odds rules based on five different online bookmakers between 77% and 79% of the matches correctly.

6 Conclusion

Collective recognition is a simple forecasting heuristic that bets on the fact that people's recognition knowledge of names of competitors is a proxy for their competitiveness. The use of the collective recognition heuristic is, of course, not limited to the domain of sports. It can be applied in virtually any domain for criteria that matter to the public and thus are likely to be reflected in people's knowledge and ignorance about the world. The Scottish historian Thomas Carlyle did "(...) not believe in the collective wisdom of individual ignorance" in political decision making. A small but growing set of data suggests that had he considered the forecasting of sport events, he might have placed more trust into the collective wisdom of individual ignorance.

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*, 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 & Exercise*, *10*, 116–123.
- Archive.org. (2008). *2008 European Championship Predictions*. Retrieved from <http://www.archive.org/details/2008EuropeanChampionshipPredictions>
- Armstrong, J. S. (2001). Combining forecasts. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 417–439). Norwell, MA: Kluwer Academic Publishers.
- Armstrong, J. S. (2005). The forecasting canon: Nine generalizations to improve forecast accuracy. *Forecasting: The International Journal of Applied Forecasting*, *1*, 29–35.
- Batchelor, R. A., & Dua, P. (1995). Forecaster diversity and the benefits of combining forecasts. *Management Science*, *41*, 68–75.
- BBC Sport. (2000). *The world rankings riddle*. Retrieved from <http://news.bbc.co.uk/sport2/hi/football/1081551.stm>
- Ben-Naim, E., Vazquez, F., & Redner, S. (2006). Parity and predictability of competitions. *Journal of Quantitative Analysis in Sports*, *2*(4/1).
- Bennis, W. M., & Pachur, T. (2006). Fast and frugal heuristics in sports. *Psychology of Sports and Exercise*, *7*, 611–629.
- Betexplorer.com. (2010a). *World Cup 2006 Germany stats, Soccer - International - tables, results*. Retrieved from <http://www.betexplorer.com/soccer/international/soccer-world-cup-germany-2006>
- Betexplorer.com. (2010b). *Euro 2008 (AUT, SUI) results & stats*. Retrieved from <http://www.betexplorer.com/soccer/international/euro-2008-aut-sui/results>
- Betexplorer.com. (2010c). *Euro 2004 Portugal stats, Soccer - International - tables, results*. Retrieved from <http://www.betexplorer.com/soccer/international/euro-2004>
- 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.
- Boulier, B. L., Stekler, H. O., & Amundson, S. (2006). Testing the efficiency of the National Football League betting market. *Applied Economics*, *38*, 279–284.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, *78*, 1–3.
- Brown, L. D., Cai, T. T., & DasGupta, A. (2002). Confidence intervals for a binomial proportion and asymptotic expansions. *Annals of Statistics*, *30*, 160–201.
- Camerer, C. F., & Johnson, E. J. (1991). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In K. A. Ericsson & J. Smith (Eds.), *Towards a general theory of expertise: Prospects and limits* (pp. 195–217). New York, NY: Cambridge Press.
- Caudill, S. B. (2003). Predicting discrete outcomes with the maximum score estimator: The case of the NCAA men's basketball tournament. *International Journal of Forecasting*, *19*, 313–317.
- Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559–583.

- Clemen, R. T., & Winkler, R. L. (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis*, *19*, 187–203.
- Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*, 571–582.
- del Corral, J., & Prieto-Rodríguez, J. (2010). Are differences in ranks good predictors for Grand Slam tennis matches? *International Journal of Forecasting*, *26*, 551–563.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In J. Kittler & F. Roli (Eds.), *First international workshop on multiple classifier systems, lecture notes in computer science* (pp. 1–15). New York, NY: Springer.
- Dixon, M., & Pope, P. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting*, *20*, 697–711.
- Evanschitzky, H., & Armstrong, J. S. (2010). Replications of forecasting research. *International Journal of Forecasting*, *26*, 4–8.
- FIFA.com (2010a). *FIFA/Coca-Cola World Ranking Schedule*. Retrieved from <http://www.fifa.com/worldfootball/ranking/procedure/men.html>
- FIFA.com. (2010b). *The FIFA/Coca-Cola World Ranking*. Retrieved from <http://www.fifa.com/worldfootball/ranking/lastranking/gender=m/fullranking.html>
- Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of the football betting market. *International Journal of Forecasting*, *21*, 551–564.
- Forrest, D., & McHale, I. (2007). Anyone for tennis (betting)? *European Journal of Finance*, *13*, 751–768.
- Franck, E., Verbeek, E., & Nüesch, S. (2010). Prediction accuracy of different market structures—bookmakers versus a betting exchange. *International Journal of Forecasting*, *26*, 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*, 73–88.
- Gambling Commission (2009). *Gambling Industry Statistics 2008/09*. Retrieved from <http://www.gamblingcommission.gov.uk>
- Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior*. New York, NY: Oxford University Press.
- Gil, R., & Levitt, S. D. (2007). Testing the efficiency of markets in the 2002 World Cup. *Journal of Prediction Markets*, *1*, 255–270.
- Goddard, J. (2005). Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, *21*, 331–340.
- Goddard, J., & Asimakopoulos, I. (2004). Forecasting football results and the efficiency of fixed-odds betting. *Journal of Forecasting*, *23*, 51–66.
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75–90.
- Goldstein, D. G., & Gigerenzer, G. (2009). Fast and frugal forecasting. *International Journal of Forecasting*, *25*, 760–772.
- Graefe, A., & Armstrong, J. S. (2009). *The popularity heuristic: Using search query data for forecasting*. Manuscript in preparation. Retrieved from <http://www.andreas-graefe.org/images/articles/popularityheuristic.pdf>
- Gröschner, C., & Raab, M. (2006). Vorhersagen im Fußball: Deskriptive und normative Aspekte von Vorhersagemodellen im Sport [Forecasting soccer: Descriptive and normative aspects of forecasting models in sports]. *Zeitschrift für Sportpsychologie*, *13*, 23–36.
- Hertwig, R., & Herzog, S. M. (2009). Fast and frugal heuristics: Tools of social rationality. *Social Cognition*, *27*, 661–698.
- Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1191–1206.
- Herzog, S. M., & Hertwig, R. (2009). The wisdom of many in one mind: Improving individual judgments with dialectical bootstrapping. *Psychological Science*, *20*, 231–237.
- Hogarth, R. M. (in press). When simple is hard to accept. In P. M. Todd, G. Gigerenzer, & The ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press.
- Hvattum, L. M., & Arntzen, H. (2010). Using ELO ratings for match result prediction in association football. *International Journal of Forecasting*, *26*, 460–470.
- Hyndman, R. J. (2010). Encouraging replication and reproducible research. *International Journal of Forecasting*, *26*, 2–3.
- Katsikopoulos, K. V. (2010). The less-is-more effect: Predictions and tests. *Judgment and Decision Making*, *5*, 244–257.
- Klaassen, F. J. G. M., & Magnus, J. R. (2003). Forecasting the winner of a tennis match. *European Journal of*

- Operational Research*, 148, 257–267.
- Lebovic, J., & Sigelman, L. (2001). The forecasting accuracy and determinants of football rankings. *International Journal of Forecasting*, 17, 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, 471–481.
- MacGregor, D. G. (2001). Decomposition for judgmental forecasting and estimation. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 107–124). Norwell, MA: Kluwer Academic.
- Makridakis, S., & Hibon, M. (1979). Accuracy of forecasting: An empirical investigation (with discussion). *Journal of the Royal Statistical Society, Series A*, 142, 97–145.
- Menschel, R. (2002). *Markets, mobs, and mayhem*. New York: Wiley.
- Pachur, T. (2010). Recognition-based inference: When is less more in the real world? *Psychonomic Bulletin & Review*, 17, 589–598.
- 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.
- Pachur, T., Bröder, A., & Marewski, J. N. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making*, 21, 183–210.
- Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 983–1002.
- Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. G. (in press). When is the recognition heuristic an adaptive tool? In P. Todd, G. Gigerenzer, & the ABC Research Group (Eds.), *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press.
- Pleskac, T. J. (2007). A signal detection analysis of the recognition heuristic. *Psychonomic Bulletin & Review*, 14, 379–391.
- Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, 19, 251–271.
- Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111–196). Cambridge, MA: Blackwell.
- Sauer, R. D. (1998). The economics of wagering markets. *Journal of Economic Literature*, 36, 2021–2064.
- Scheibehenne, B., & Bröder, A. (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition? *International Journal of Forecasting*, 23, 415–426.
- Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, 112, 610–628.
- Serwe, S., & Frings, C. (2006). Who will win Wimbledon 2003? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making*, 19, 321–332.
- Smith, T., & Schwertman, N. C. (1999). Can the NCAA basketball tournament seeding be used to predict margin of victory? *American Statistician*, 53, 94–98.
- Smithson, M. (2010). When less is more in the recognition heuristic. *Judgment and Decision Making*, 5, 230–243.
- Stefani, R. T. (1980). Improved least squares football, basketball, and soccer predictions. *IEEE Transactions on Systems, Man, and Cybernetics*, 10, 116–123.
- Stefani, R. T. (1997). Survey of the major world sports rating systems. *Journal of Applied Statistics*, 24, 635–646.
- 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.
- 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, 426–444.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p-values. *Psychonomic Bulletin & Review*, 14, 779–804.
- Wilcox, R. R. (2005). *Introduction to robust estimation and hypothesis testing* (2nd ed.). San Diego, CA: Elsevier Academic Press.
- Wilcox, R. R. (n.d.). *Rallfun-v11* [Statistical functions for R]. Retrieved from <http://www-rcf.usc.edu/~rwilcox/Rallfun-v11>
- Winkler, R. L. (1971). Probabilistic prediction: Some experimental results. *Journal of the American Statistical Association*, 66, 675–685.
- Yates, J. F. (1982). External correspondence: Decompositions of the mean probability score. *Organizational Behavior and Human Performance*, 30, 132–156.
- Yates, J. F. (1994). Subjective probability accuracy analysis. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 381–410). Chichester, England: Wiley.