Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Acta Psychologica 125 (2007) 99–116
acta psychologica
www.elsevier.com/locate/actpsy

Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events

Thorsten Pachur *, Guido Biele

Max Planck Institute for Human Development, Center for Adaptive Behavior and Cognition, Lentzeallee 94, 14195 Berlin, Germany

Received 7 November 2005; received in revised form 16 June 2006; accepted 3 July 2006
Available online 10 August 2006

Abstract

Whereas previous studies on how people make forecasts of sports events focused primarily on experts, we examined how laypeople do this task. In particular, we (a) tested the recognition heuristic [Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review, 109*, 75–90], which requires partial ignorance, against four alternative mechanisms in describing laypeople's forecasts for the European Soccer Championships 2004; (b) evaluated how well recognition predicted the outcomes of the matches compared to direct indicators of team strength (e.g., past performance, rankings); and (c) studied the *less-is-more effect*—the phenomenon that knowing less leads to more correct forecasts than knowing more—which can occur when the recognition heuristic is used. Two groups of participants (laypeople, experts) made forecasts for the first-round matches of the tournament. Of the five candidate mechanisms, the recognition heuristic predicted laypeople's forecasts best: when applicable, it accounted for 90% of the forecasts. The recognition heuristic correctly predicted the actual winner of the matches substantially better than chance but did not achieve the accuracy of direct indicators of team strength. The experts made more correct forecasts than the laypeople. Moreover, we found no benefit of ignorance among the group of laypeople, although the conditions for a less-is-more effect specified by Goldstein and Gigerenzer were fulfilled.

PsycINFO classification: 2340; 3720

Keywords: Forecasting; Heuristics; Sports

* Corresponding author. Tel.: +49 30 82406 691; fax: +49 30 82406 394.
E-mail address: pachur@mpib-berlin.mpg.de (T. Pachur).

0001-6918/$ - see front matter
doi:10.1016/j.actpsy.2006.07.002

1. Introduction

Panem et circenses ("bread and circuses")—thus did the poet Juvenal describe the efficient formula used by the Roman emperors to keep the population peaceful. Juvenal's comment, however, not only describes a political strategy of its time: It also reflects the early importance of sports games in society. Yet in ancient Rome it was not watching the gladiator games and chariot races alone that amused the people—predicting and betting on the outcome of the games were equally important (e.g., Weeber, 1998). The close link between sports and prediction continues today. In this article, we are interested in people's forecasts for one contemporary equivalent of the Roman games, soccer.

Until recently, research on the psychological mechanisms underlying sports forecasting has been primarily focused on the forecasts of experts (Boulier & Stekler, 2003; Cantinotti, Ladouceur, & Jacques, 2004; Forrest & Simmons, 2000; Heath & Gonzalez, 1995; Kaplan, 1980; Koehler, 1996; Ladouceur, Giroux, & Jacques, 1998; Vertinsky, Kanetkar, Vertinsky, & Wilson, 1986; for exceptions see, e.g., Gilovich, Vallone, & Tversky, 1985; Heit, Price, & Bower, 1994). As a consequence, relatively little is known about how laypeople—who have neither extensive domain-specific knowledge nor experience with elaborate judgment strategies—make forecasts for sports events. This apparent neglect is surprising, as millions of non-experts engage in sports forecasting every week.[1] The focus of the present study was to investigate how laypeople make forecasts about the winners in sports competitions, and we took advantage of the 2004 European Soccer Championships (EURO 2004) to study this issue.

Specifically, we tested one recently proposed model of lay judgment, the recognition heuristic (Goldstein & Gigerenzer, 2002), and, unlike in previous tests of the heuristic, compared it with other candidate mechanisms for describing laypeople's forecasts. Second, we evaluated the predictive value of recognition as a cue against a number of direct indicators of team strength (rankings, recent performance, odds). Third, we attempted to trace one possible consequence of the recognition heuristic, the less-is-more effect, which refers to the phenomenon that knowing fewer objects in a domain can lead to higher forecasting accuracy than knowing more.

The remainder of this article is organized as follows. We first provide a description of the recognition heuristic, give a short summary of existing applications of the heuristic to sports forecasting, and explain how the use of the heuristic can lead to a counterintuitive benefit of limited knowledge. We then describe four alternative mechanisms for lay prediction and define benchmark cues that we used to evaluate recognition in terms of its ability to predict the outcome of sports events. Finally, we report a study on forecasts of matches at the EURO 2004 in which we tested the candidate mechanisms, mapped the boundaries of recognition's predictive strength, and investigated to what degree partial ignorance was beneficial for making accurate forecasts.

2. Forecasting by exploiting partial and systematic ignorance: the recognition heuristic

One conclusion from research on sports forecasting by experts is that although experts are able to make accurate forecasts to some degree, they deviate from optimality

[1] It is estimated that annually up to $380 billion is spent on sports betting (Macy, 1999), and in 2003, almost every fifth adult American had engaged in legal or illegal sports betting (American Gaming Association, 2004). In light of the sheer number of bets, a substantial proportion of bettors will have only superficial sports knowledge and bet only occasionally and mainly for hedonistic reasons.

substantially. Specifically, experts seem to weight information about the competing sports actors both inaccurately and inconsistently (e.g., Boulier & Stekler, 2003; Cantinotti et al., 2004; Forrest & Simmons, 2000). Laypeople have quite a different problem. Not only do they usually not have much information about the competing sports actors (let alone know how to weight the information), they sometimes have not even heard of them. How are forecasts made under such adverse circumstances?

Goldstein and Gigerenzer (2002) recently proposed a model of heuristic decision making that seems particularly appropriate for modeling lay prediction, and which has also been applied in the context of sports forecasting (Andersson, Ekman, & Edman, 2003; Ayton & Önkal, submitted for publication; Serwe & Frings, in press; Snook & Cullen, 2006). The *recognition heuristic* not only tolerates but even requires incomplete knowledge. Specifically, applied to the task of predicting the winning team of a soccer match, the heuristic exploits the situation when one has heard of only one of the teams in a competition.[2] The recognition heuristic works as follows: "If only one of the teams is recognized, predict that the recognized one will win."

Obviously, the recognition heuristic will not always apply (see also Section 7), nor will it always make a correct forecast. It requires (a) partial ignorance to make a forecast in the first place and (b) systematic ignorance to make a correct forecast. If all (or none of the) teams that compete with each other are recognized, the recognition heuristic will not make a prediction. Moreover, whether a team is recognized or not must be systematic (rather than random) in the sense that recognized teams tend to be successful. As successful teams are probably more often talked about, the teams one has heard of will also, overall, be successful in the future. So arguably, reliance on recognition will often lead to correct forecasts in the sports domain. In general, it is said that in such an environment the heuristic is "ecologically" rational.

Previously, the heuristic has been intensely studied in inference tasks, such as deciding which of two cities is larger. The application of the recognition heuristic to sports forecasting provides an important test case for the heuristic. First, it obviates the objection to previous tests of the recognition heuristic that recognition was confounded with criterion knowledge (Oppenheimer, 2003), and thus the heuristic was not tested adequately. In forecasts, which refer to a future state, such a confounding is excluded. Second, the domains in which the heuristic has been primarily studied previously (e.g., city size) were static ones—that is, the criterion values of the objects remain relatively constant. The sports domain, by contrast, is a dynamic one, with the strengths of the competitors varying considerably over time (cf. Serwe & Frings, in press). As a consequence, a competitor might be known for past success, and in such a case recognition can be systematically misleading.

Some recent work suggests that laypeople actually use the recognition heuristic in the sports domain. For instance, in Serwe and Frings (in press), tennis laypeople made forecasts for the matches at the 2003 Wimbledon tournament and also indicated which players they had heard of before. In more than 90% of the matches in which a recognized player played against an unrecognized one, the participants made the forecast that the recognized player would win. Similarly, Snook and Cullen (2006) found that when participants were asked to infer which of two NHL players had a higher number of career points,

[2] For a discussion of the difference between recognition and related concepts such as availability and fluency, see Goldstein and Gigerenzer (2002), Pachur et al. (in press), and Schooler and Hertwig (2005).

a recognized player was selected over an unrecognized player 96% of the time. (Note, however, that this was not a forecasting task.)[3]

Snook and Cullen (2006) also showed that recognition was a very useful piece of information. When a recognized player was inferred to have more career points than an unrecognized one, this inference was correct 81–94% of the time. Recognition was a good predictor of success also in Serwe and Frings's (in press) study on Wimbledon tennis matches, where a recognized player won against an unrecognized one almost 70% of the time. In addition, Serwe and Frings made the surprising observation that recognition fared extremely well when compared to other, tennis-specific cues. In fact, recognition predicted the actual winner better than the Association of Tennis Professionals (ATP) rankings, which are based on a large amount of information about the players' previous performance. We will perform a similar comparison in our study to see whether this astonishing result can be replicated.

Goldstein and Gigerenzer (2002) demonstrated an interesting phenomenon that can arise from the use of the recognition heuristic (when recognition is a good predictor). The achievable accuracy based on the recognition heuristic in comparisons where it can be applied is expressed by the *recognition validity*, indexed by α. For instance, in the context of forecasting soccer matches, the recognition validity is calculated as the number of matches in which a recognized team plays against an unrecognized one and wins, divided by the number of matches in which only one team is recognized. Moreover, the proportion of correct forecasts when both teams are recognized (in which case the recognition heuristic cannot be applied) is the *knowledge validity*, indexed by β. Goldstein and Gigerenzer showed that when $\alpha > \beta$, a person who recognizes fewer teams can achieve a higher (or equal) overall forecasting accuracy than a person who recognizes relatively more teams, a phenomenon Goldstein and Gigerenzer labeled the *less-is-more effect*. For illustration, when none of the teams are recognized, one has to guess, and only chance accuracy is achieved. When more and more teams are recognized, overall accuracy increases, but when more than about half of the teams are recognized the number of instances in which the recognition heuristic can be applied decreases and the higher α contributes less to the overall accuracy (which is then increasingly determined by the lower β; for details see Goldstein & Gigerenzer, 2002). As a consequence, accuracy decreases.

Can such an effect be observed empirically? There are some indications that less knowledge does not necessarily lead to less accurate—and sometimes even leads to more accurate—predictions. Andersson, Edman, and Ekman (2005), for instance, compared expert and lay predictions of the first-round outcome of the 2002 Soccer World Championships and found that the lay participants were more accurate than the experts. (Moreover, the lay participants seemed to use their lack of knowledge of some of the teams for making a forecast.) In a study on forecasts of English soccer matches, Ayton and Önkal (submitted for publication) observed that Turkish participants—though they probably knew very little about English soccer teams (but might have heard of some of the city names)—performed

[3] Ayton and Önkal (submitted for publication) and Andersson et al. (2003) also examined (and confirmed) people's use of their lack of knowledge for making forecasts. However, in these studies recognition was defined differently from how it was originally conceptualized by Goldstein and Gigerenzer (2002). Moreover, the forecasting task in Andersson et al. was a multiple alternative, rather than a paired comparison task, for which the recognition heuristic has been described. For these reasons, these studies are not considered as tests of the recognition heuristic here.

nearly as accurately as knowledgeable British participants. In a more refined analysis, Snook and Cullen (2006) obtained a similar result also across lay participants differing in the number of sports actors they recognized: Participants who recognized almost all players were not better than participants who recognized only about half of the players. Currently, the study by Snook and Cullen seems to be the only attempt to trace a less-is-more effect within a group of participants (see, however, Reimer & Katsikopoulos, 2004 for a demonstration between groups).

To conclude, the current evidence provides some support for the notion that people with limited sports-specific knowledge use their lack of recognition—that they have heard of only one of two competitors—to make forecasts of the outcomes of sports events. A major drawback of these (and other) applications, however, is that none of them tested the recognition heuristic against alternative mechanisms. For instance, people could base their judgments not on recognition but on a cue that is correlated with recognition, in which case the heuristic could still predict the data well. Putting the recognition heuristic to such a test is one goal of our current study. Second, recognition seems to be a rather useful predictor of the outcome of sports events, with some evidence suggesting that it can equal or even exceed the accuracy of highly sophisticated predictors that summarize a large amount of information. Testing the robustness of this result is our second goal. Finally, it has been observed that in judgment tasks in which recognition was relied on, more knowledge was not associated with higher judgmental accuracy (and was sometimes even associated with lower accuracy), akin to the less-is-more effect described by Goldstein and Gigerenzer (2002). Our third goal was to examine whether this phenomenon can be replicated.

3. Hypotheses and research questions

Against this background, the following hypotheses guided our application of the recognition heuristic to the EURO 2004:

- The recognition heuristic predicts laypeople's forecasts better than other plausible candidate mechanisms (specified below).
- Recognition predicts the actual winners of the matches better than direct indicators of team strength, such as rankings and recent performance.

In addition, we sought to obtain an answer to the following question:

- Can we find indications for a benefit of limited knowledge (i.e., less-is-more effect)? Specifically, do laypeople make more correct forecasts than experts? Do laypeople who recognize about half of the teams make more accurate forecasts than laypeople who recognize all the teams?

4. Overview of the study

Before the tournament took place, we asked groups with different degrees of knowledge (laypeople and experts) to forecast the winners of the 24 first-round matches of the EURO 2004 and to indicate for each participating country whether they had heard of its national soccer team. Use of the recognition heuristic was tested by looking at how often the recognized teams, when playing against an unrecognized team, were judged to be more likely to win. The performance of the recognition heuristic in describing participants' forecasts was compared to alternative mechanisms of lay prediction. To assess the predictive accuracy of the participants and recognition, we compared the predictions with the actual outcomes of the matches. To examine the potential benefit of limited knowledge, we used two approaches: First, we explored—between groups—whether the less knowledgeable laypeople would make better forecasts than the knowledgeable experts. Second, we explored—within the group of laypeople—whether individuals who recognized fewer teams would make better forecasts than individuals who recognized more teams.

Before we describe alternative descriptive models of lay prediction, let us examine whether recognition could be useful in forecasting the winners of the matches of the EURO 2004. The central assumption here is that strong teams tend to be mentioned in the media more frequently than weak ones, and that more frequently mentioned teams, in turn, are more likely to be recognized. Consistent with this assumption, we found that presence in the media (defined as the number of co-occurrence of the country names with the term "soccer" in the German print media, using the COSMAS II word corpus[4]) correlated highly with the Fédération Internationale de Football Association (FIFA) ranks of the teams (a plausible indicator of team strength), $r_s = -0.77$. Furthermore, using the recognition rates collected in our study (see below), the mentioned frequency and the proportion of lay participants recognizing the teams (*collective recognition*; Goldstein & Gigerenzer, 2002) were highly correlated, $r_s = 0.94$. Finally, there was also a strong correlation between recognition and FIFA ranks, $r_s = -0.81$, illustrating the ecological rationality of the recognition heuristic for judging a team's strength.

4.1. Alternative mechanisms of lay prediction

As discussed above, a central problem of previous studies on the recognition heuristic is that, as a model of people's judgments, the heuristic was not tested against alternative models. Here we contrast the recognition heuristic with four alternative accounts of how people with limited soccer knowledge might make forecasts about the winners of the matches at the EURO 2004.

What are other plausible cues laypeople might use to forecast the winner of a soccer match with national teams? Although laypeople have by definition only little soccer-specific knowledge about the teams (e.g., their recent performance, current players) they will probably know something about the countries of the participating teams (e.g., population size, economic strength). Since such general knowledge might also be probabilistically related to success in sports, it could be used to judge which team is more likely to win. In the following, we describe four alternative mechanisms that use country knowledge as cues to make a forecast. Three of the mechanisms make a forecast based on only one cue (as does the recognition heuristic); the last is a simple additive mechanism that integrates multiple cues.

[4] COSMAS (Corpus Search, Management and Analysis System) is the largest online archive of German literature (e.g., encyclopedias, books, and newspaper articles) and can be accessed on http://www.ids-mannheim.de/cosmas2/. At the time of our analysis, the corpus contained around 1.9 billion words. For our search, a co-occurrence was defined as the country name and the term "soccer" occurring together within a maximum span of 10 words.

4.1.1. Gross domestic product (GDP)

The rationale of this mechanism is that richer countries can be expected to invest more money in the promotion of talented soccer players than poorer countries. Therefore, according to GDP the team from the country with the higher per capita gross domestic product (as an indicator of economic strength of the country) is judged to be more likely to win. (To generate the predictions of GDP and the following mechanisms, information about the countries was obtained from http://www.nationmaster.com/index.php.)

4.1.2. Population size of the country (POP)

All things being equal, countries with a large population will produce a larger number of outstanding players than countries with a small population. Based on this rationale, according to POP, the country with the larger population is judged to be more likely to win.

4.1.3. Former membership in east block (EAST–WEST)

As most formerly East Block countries have joined the European Community only relatively recently, many people in Germany (where our study was run) will probably not have heard of the national soccer teams of these countries. Furthermore, many of the formerly East Block and now European countries are still relatively weak economically and have relatively small populations (the exception being Russia). Therefore, a way to implicitly combine the recognition and the GDP cues is to use former membership in the East Block as a cue. According to EAST–WEST, when a Western European team plays against a team from a country that belonged to the former East Block, the Western European team is judged to be more likely to win.

That people rely on only one cue when making a forecast is certainly a strong claim. Therefore, we also included a mechanism that integrates the three cues used by the previous mechanisms.

4.1.4. Tallying (TALLY)

According to this mechanism, the three cues—GDP, population size, and former membership in the East Block—are equally weighted and the positive evidence for each of the teams is summed. The team that has a larger sum of positive evidence is judged to be more likely to win (e.g., Dawes, 1979). (Because TALLY integrates three cues, the maximum sum is 3.)

When testing these mechanisms, it was not expected that participants would have precise knowledge of the countries' economic strengths and population sizes. Rather, we assumed a difference threshold δ, according to which the values of two countries a and b on cue x are perceived as different only if they differ by at least 40% of the larger value ($\delta = \max\{x_a, x_b\} \times 0.4$). This threshold was determined as follows. We asked a group of participants to compare pair-wise the 16 participating countries (120 pair comparisons) in terms of either population size ($n = 20$) or GDP ($n = 20$). Mean accuracies were 81.9% (SD = 6.2) and 87.3% (SD = 3.7) for population size and GDP, respectively. In both conditions, two countries could be reliably discriminated (i.e., accuracy >50%) as soon as their population sizes or GDPs, respectively, differed by at least 40% of the larger of the two values. When in testing the mechanisms the difference does not exceed this threshold, the cue does not discriminate between the two countries and in the case of GDP and POP, no prediction is made.

4.2. Direct indicators of team strength as a benchmark for recognition

How well can recognition predict success in sports? Previous studies have identified rankings (Boulier & Stekler, 1999; Serwe & Frings, in press), past performance (Boulier & Stekler, 2003; Forrest & Simmons, 2000), and betting odds (Boulier & Stekler, 2003; Serwe & Frings, in press) as robust predictors of success in sports. We chose these direct indicators of team strength to obtain benchmark levels of performance against which recognition would be compared. In addition, we also used the accuracy of experts *as a group* as a benchmark. According to the Condorcet jury theorem (Condorcet, 1785/1994), the pooled judgments of a group of individuals can be more accurate than that of the average individual (Grofman & Owen, 1986), and Forrest and Simmons (2000) showed such an effect in the sports domain as well.

4.2.1. FIFA ranking (FIFA-RANK)

The FIFA rankings are based on a large amount of information concerning the performance of the teams in the last 8 years, such as number of wins and goals, whether the match was at home or away, importance of the match, regional strength, and period (with less recent performance given progressively less weighting). We tested how often the team with the higher FIFA rank (before the tournament) actually won. The rankings were obtained from www.fifa.com.

4.2.2. Performance in qualifying round (QUALY)

Teams were ranked in accordance with their performance in the qualifying round for the EURO 2004 (there were no ties). Relevant criteria were points won (three points for a win, one point for a draw) and the goals scored and conceded. Because Portugal, as host team, qualified automatically, the matches that included Portugal were not included in the test of this cue. We tested how often the team with the higher rank according to performance in the qualifying round won.

4.2.3. Betting odds (BET)

We took the betting odds from online bettors[5] and determined how often the outcome with the lower odds (odds for draws were excluded) actually occurred. It should be noted that for two reasons, the strength of the odds for predicting success should not be too surprising. First, odds represent an aggregate of predictions by a large number of people (expressed through their betting behavior). Second, in contrast to the other indicators, odds are continually updated, so they contain information about the course of the tournament.

4.2.4. Expert majority (MAJ-Exp)

We determined for every match the team that the majority of experts had selected to be more likely to win, and then checked how often this team actually won. One match (The Netherlands vs. Czech Republic) was not included in the analysis as an equal number of experts had selected each team.

[5] Odds were obtained (after the tournament) from BetExplorer (http://www.betexplorer.com/soccer/international/euro-2004/league.php?group = 0&lastXMatches = 9999&round = 1), which provides a summary of the final offers of around 70 online betting companies.

5. Method

5.1. Participants

To obtain forecasts by laypeople, we recruited 121 (62 female and 59 male; mean age 29.9 years, range 11–72) participants at various public places in Berlin (cafeterias, museums). Soccer experts were recruited by contacting (by email) editors of sports sections at major German newspapers as well as TV and radio stations.[6] They were also asked to forward the announcement to colleagues. Twenty soccer experts (2 female and 18 male, mean age 39.5 years, range 28–65) participated. Consistent with their professional status, the experts indicated (on a 7-point scale) higher levels of both self-reported interest in soccer and soccer knowledge than the lay participants (knowledge: $M = 6.8$, $SD = 0.4$ vs. $M = 3.02$, $SD = 1.87$; $t(130.0) = 19.65$, $p = .001$, Cohen's $d = 1.88$; interest: $M = 6.1$, $SD = 0.8$ vs. $M = 2.79$, $SD = 1.74$; $t(59.1) = 14.08$, $p = .001$, $d = 2.31$).

5.2. Tasks

5.2.1. Forecasting task

Participants were presented with a randomly ordered list of the 24 matches of the first round of the EURO 2004 (both the positions of the teams and the order of the matches were determined randomly) and asked to indicate which team they thought was more likely to win each match. As the recognition heuristic, the focus of our investigation, is formulated for a two-alternative forced choice task, a draw was not a response option. In addition to the forecasting task, participants completed a task in which they ranked each of the 16 national teams such that the rank given to a team would denote that the team was more likely to win against all teams with a lower rank. The responses in this ranking task were used to test the robustness of some of the results of the forecasts for the first round (see Footnote 9).

5.2.2. Recognition task

Participants were presented with the names of the countries in alphabetical order and asked to indicate whether they had heard or read about the national soccer team of the country before participating in our study.

5.3. Procedure

The data collection took place in early June 2004, within the 14 days preceding the EURO 2004. The questionnaires were administered to the laypeople as a paper and pencil task; the experts completed an online version. After the participants had indicated their age, sex, and profession, they completed the forecasting and recognition tasks in this order. As previous studies found no indication of an effect of task order (e.g., Goldstein & Gigerenzer, 2002; Pachur et al., in press), we decided to use the same task order for all participants. We put forecasting before recognition because the reverse could have artificially increased participants' attention to whether they recognized a team or not. All participants

[6] Note that we thus define sports experts according to their (expected) amount of sports knowledge (cf. Phillips, Klein, & Sieck, 2004), rather than a priori according to their forecasting proficiency (cf. Camerer & Johnson, 1991). Given that the online questionnaire was in German, we assume that all participants were German speaking.

were offered a monetary incentive for the forecasting task: They were informed that the participants with first, second, and third best accuracy would receive €30, €20, and €10 (US$36.6, US$24.4, and US$12.2), respectively. Whereas we assumed that experts would be intrinsically motivated to participate, laypeople received a chocolate bar as compensation. The laypeople took around 15 min to complete the questionnaire (we have no reason to assume that the experts had different completion times).

6. Results

One lay participant gave extremely irregular responses (i.e., deviated, on crucial variables such as forecasting accuracy, α, and β, by more then two standard deviations from the mean) and was excluded from the analysis.

6.1. How well did the recognition heuristic predict laypeople's forecasts?

On average, the lay participants recognized 11.1 (SD = 3.4, range 3–16) of the 16 teams (69.2%). Overall, they were able to use the recognition heuristic for, on average, 9.22 (SD = 4.93) of the 24 matches (38.4%). Of the 20 experts, in contrast, only 2 did not recognize all the teams and could use the recognition heuristic 7 and 10 times, respectively. Knowing too much, the other 18 experts could never use the recognition heuristic.

To examine how well the recognition heuristic could predict lay participants' forecasts, we looked at all matches where a participant had heard of one of the teams but not the other, irrespective of the outcome of the match. We then checked how often the recognized team was picked as more likely to win. Across the 103 lay participants who could apply the recognition heuristic at least once, the mean percentage of forecasts in line with the recognition heuristic was 90.5% (SD = 12.2; median and mode were 100%; see Table 1). The two experts who had the necessary ignorance to be able to use the recognition heuristic picked the recognized team in 70% and 71.4% of the cases.

6.2. How well did the other candidate mechanisms predict laypeople's forecasts?

To test how well the alternative mechanisms, GDP, POP, EAST–WEST, and TALLY, could predict which team our lay participants picked as more likely to win, we determined

Table 1
Results for the first-round matches

	Laypeople	Experts
Percentage of correct forecasts (*M*)	64.5[a]	76.6[a]
SD	10.6	9.3
Percentage of forecasts in line with RH (*M*)	90.5	70.7[b]
SD	12.2	1.01
Recognition validity α (*M*)	0.71[a]	0.68[a,b]
SD	0.18	0.25
Knowledge validity β (*M*)	0.60[a]	0.77[a]
SD	0.24	0.10

RH—recognition heuristic.

[a] The eight matches that ended in a draw are not included in the analysis.

[b] Results based on $n = 2$.

Table 2
How well did the five different mechanisms predict lay participants' forecasts? Results are given for cases in which a participant recognized one team but not the other

	GDP	POP	EAST–WEST	TALLY	RH
Proportion of correctly predicted forecasts (*M*)	0.75 (0.70)	0.80 (0.73)	0.81 (0.75)	0.85 (0.77)	0.91 (0.90)
SD	0.20 (0.13)	0.20 (0.11)	0.20 (0.14)	0.14 (0.11)	0.12 (0.12)
Applicability (*M*)	0.26 (0.58)	0.29 (0.71)	0.24 (0.54)	0.33 (0.83)	0.38 (0.54)

The values for all cases in which the individual mechanisms made an unambiguous prediction are given in brackets. For the recognition heuristic (RH), the values in brackets are those when collective recognition is used. See text for description of the mechanisms. A mechanism's applicability is defined as the proportion of matches for which it makes an unambiguous prediction.

for each mechanism, when it made an unambiguous prediction, how often this prediction coincided with the lay participants' forecasts. Table 2 reports the proportion of correctly predicted forecasts (averaged across participants) for the subset of cases where the recognition heuristic made a prediction (to allow for a direct comparison with the recognition heuristic), as well as the proportion of matches where the mechanism made an unambiguous prediction (i.e., the mechanisms' applicabilities). Statistical tests were based on these cases. The results are basically the same when all cases where the mechanisms make a prediction are considered (the corresponding values are reported in Table 2 in brackets).

Compared to the alternative mechanisms the recognition heuristic correctly predicted the largest proportion of forecasts, and sign tests show that it predicted the forecasts better for more participants than any of the other mechanisms (exact test; all ps = 0.0001).[7] Of the four alternative mechanisms, TALLY and EAST–WEST predicted laypeople's forecasts best (with 84.7% and 80.7% correctly predicted forecasts, respectively) but lagged considerably behind the 90.5% achieved by the recognition heuristic.

A possible objection to the preceding analysis is that the recognition heuristic might have had an edge over the other mechanisms as it could, for a given match, make different predictions for different participants, whereas the other mechanisms made the same prediction for each participant. To deal with this objection, we also tested a version of the recognition heuristic that made the same prediction for each participant. We used collective recognition (Goldstein & Gigerenzer, 2002) for this purpose. It was thus predicted that the team that more participants had heard of would be picked (to ensure comparability with the other mechanisms, again a threshold of 40% of the larger value was assumed). It turned out that even with this stricter test, the recognition heuristic predicted participants' forecasts best. On average 89.5% (SD = 11.5) of lay participants' forecasts were in line with the (collective) recognition heuristic (applicability = 0.54), and it correctly predicted the forecasts better for a larger number of participants than the second-best mechanism, TALLY (on all cases where the mechanisms made a prediction; sign test, $p = .0001$). Thus, the better fit of the

[7] As the pair-wise (across participants) differences between the proportions of forecasts correctly predicted by the five mechanisms violated normality for five of the 10 pairs of mechanisms, the mechanisms were compared based on sign tests (exact tests).

recognition heuristic was not due to its predictions being tailored to the individual participants.[8]

To examine one possible reason why the recognition heuristic might have been preferred as a forecasting strategy, we calculated in addition how well the five candidate mechanisms, when they made an unambiguous prediction, predicted the actual outcomes of the matches (for this analysis, no threshold was assumed). Eight of the 24 matches of the first round ended in draws and these matches were not included, as a draw was not modeled by the mechanisms (nor was a draw a response option in the participants' forecasting task). Recognition (0.71) and EAST–WEST (0.70) turned out to have the highest validities, defined as the proportion of matches for which the winner was predicted correctly (they did not differ significantly from each other: one-sample t-test, $t(103) = 0.50$, $p = .62$; $d = .007$). The recognition validity was significantly higher than the validity of TALLY (0.67), which had the next highest validity ($t(103) = 2.22$, $p = .03$; $d = .22$). The validities of GDP and POP were 0.50 and 0.63, respectively. Overall, this suggests that the recognition heuristic's validity might have been one factor fostering its use over the other mechanisms.

6.3. Testing the limits of recognition: a comparison with direct indicators of team strength

We contrasted the ability of recognition to correctly predict the winners of the soccer matches with the predictive strength of four direct indicators of team strength: rankings (FIFA-RANK), past performance (QUALY), betting odds (BET), and expert majority (MAJ-Exp). To facilitate a direct comparison with recognition, we determined, in all cases in which the recognition heuristic made a prediction, the proportion of matches for which the four direct indicators of team strength correctly predicted the winner (excluding draws). Fig. 1 shows the results (averaged across participants). In contrast to the results reported by Serwe and Frings (in press), in achieving 63% correct predictions recognition clearly falls behind its competitors. Overall, the predictions based on the majority of experts (MAJ-Exp: 87%) and the teams' performance in the qualifying round (QUALY: 85%) achieved the highest accuracy.

The pattern also holds when collective recognition is used and all cases are considered in which the indicators make a prediction, as shown in the black bars in Fig. 1. As an aside, the two best direct indicators after MAJ-Exp, FIFA-RANK and QUALY, outperformed the mean accuracy of individual experts (76.6%, is indicated by the dotted line). In addition, the experts *as a group* (MAJ-Exp) outperformed 16 of the 20 *individual* experts.

6.4. Was limited knowledge beneficial?

As mentioned above, the experts and lay participants differed substantially in terms of the number of teams they had heard of before. Moreover, there was also substantial variation within the group of laypeople. For instance, 17 lay participants had heard of fewer

[8] As could be expected given the high correlation between recognition and the FIFA ranks shown earlier, the teams that lay participants judged to be more likely to win were often those with the better FIFA rank (82.5%). The recognition heuristic, both when based on individual and on collective recognition, was still the better predictor of the forecasts (sign tests (exact): $ps = .0001$). In general, given lay participants' relatively low forecasting accuracy and the very good performance of the FIFA ranks as predictor of the actual winner (see text), it is unlikely that the lay participants actually used knowledge about the FIFA ranks.

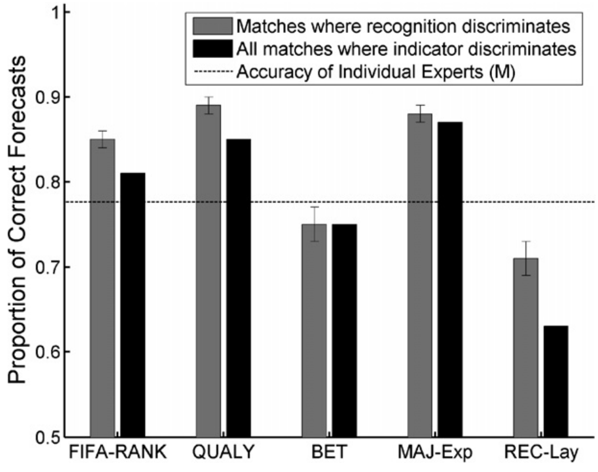

Fig. 1. Mean proportion of correct predictions (across participants) of the direct indicators of team strength when one team was recognized but not the other. The bars indicate standard errors. REC-Lay is recognition heuristic based on the lay participants' recognition judgments. The experts' average individual forecasting accuracies are indicated by the dotted line.

than half of the 16 participating teams, and 17 participants had heard of all teams. We were interested in two possible manifestations of a benefit of limited knowledge: between groups (experts vs. laypeople) and across laypeople spanning different levels of knowledge.

6.4.1. Laypeople and experts

Comparing the lay participants' forecasting accuracy to the experts', the former did not benefit from their limited knowledge. For the 16 non-draw matches, the lay participants made on average 64.7% correct forecasts (SD = 10.3; range 37.5–85.5%), whereas the experts achieved on average 76.6% correct forecasts (SD = 9.27; range 56.3–93.8; $t(138) = 4.8$, $p = .001$, $d = 1.16$).[9] As Table 1 shows, the experts' knowledge validity β—the main determinant of their overall accuracy given that they knew practically all teams and thus could not use the recognition heuristic—was substantially higher than both the laypeople's average β, $t(68.2) = 3.01$, $p = .001$, $d = 0.73$, and their α, $t(46.7) = 1.97$, $p = .05$; $d = 0.33$. In other words, the condition specified by Goldstein and Gigerenzer (2002) for a less-is-more effect was not fulfilled.

6.4.2. Less-is-more across laypeople

This was different within the group of laypeople: Here the average α (0.71) was larger than the average β (0.60). Did this lead to a less-is-more effect? We analyzed, as a function of the number of recognized teams, the accuracy of those 62 lay participants whose

[9] Similar results also held for the matches of the knock-out phase (seven matches in total). We derived predictions for these matches from our participants' rankings (see Section 5). The mean consistency between predictions and rankings for the first round was, on average, 83.9% (SD = 12.2) for laypeople and 87.9% (SD = 7.8) for experts. Though both experts and laypeople predicted much less successfully than for the first round, the experts (47.9% correct, SD = 10.7) still predicted considerably better than the laypeople (39.3% correct, SD = 15.6), $t(139) = 2.36$, $p = .02$.

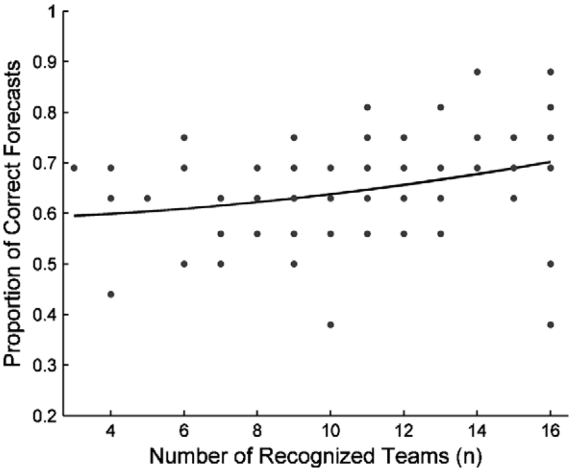

Fig. 2. Forecasting accuracy as a function of the number of recognized teams for participants whose forecasts were in line with the recognition heuristic at least 90% of the time. Note that some of the points represent multiple participants, who reached at a given level of n, the same accuracy.

forecasts were in line with the recognition heuristic at least 90% of the time (when applicable) and of those 17 lay participants who recognized all teams (for a similar analysis, see Snook & Cullen, 2006). Again the matches that ended in draws were not included in the analysis. As can be seen from Fig. 2, although on average $\alpha > \beta$, there was no indication of a benefit of ignorance: As the number of teams that participants recognized increased, there was a corresponding increase in forecasting accuracy ($r = .34$, $p = .008$).

Why was less not more? A look at the relationships between the number of recognized teams, n, and the recognition validity α and knowledge validity β reveals one possible reason. Specifically, in contrast to Goldstein and Gigerenzer's analysis, both α and β were positively correlated with n ($r = .19$, ($p = .05$) and $r = .22$ ($p = .02$), respectively). Although the correlations are not very high, note that they counteract a negative association between accuracy and n at high levels of n (implicated by a less-is-more effect). For illustration, since α and β are the main determinants of the overall accuracy (cf. Goldstein & Gigerenzer, 2002), a systematic trend toward higher values of α and β when more teams are recognized (i.e., when n is increasing), means that with increasing n the overall forecasting accuracy increases as well—the opposite of a less-is-more pattern.

7. Discussion

We compared the recognition heuristic as a model of how people with limited knowledge make forecasts of soccer matches to a set of alternative mechanisms. Moreover, we tested the heuristic's ability to predict the outcome of the events against various direct indicators of team strength. Finally, we examined the possibility that limited knowledge can be beneficial. As hypothesized, the recognition heuristic predicted the lay participants' forecasts better than the other candidate mechanisms. Our second hypothesis, however, was not supported. Although the recognition heuristic was able to predict the outcomes of the matches well above chance level, it lagged behind the direct indicators. Concerning the potential benefit of limited knowledge, there was no less-is-more effect between the lay and expert groups, as experts' knowledge validity exceeded both laypeople's recognition validity and knowledge validity. Although the conditions for a less-is-more effect seemed fulfilled for the laypeople, there was no indication of a benefit of limited knowledge within the group of lay participants either. In the following, we discuss the implications of our findings.

7.1. People's use of the recognition heuristic

Our study corroborates previous findings (Ayton & Önkal, submitted for publication; Serwe & Frings, in press; Snook & Cullen, 2006) that in forecasts of the winner in a sports competition, a recognized competitor is very often judged to be stronger than an unrecognized one. We showed that even when compared to alternative accounts of lay prediction, the recognition heuristic did best in predicting people's forecasts. It could be argued, however, that the alternative lay mechanisms were at a disadvantage as we only assumed, but did not certify, that laypeople had correct knowledge of the relevant information (GDP, population size, etc.). Although it is indeed possible that some participants did not know the correct rank order of the countries on the relevant dimensions, by using an empirically derived difference threshold (40%), we made reasonable assumptions concerning the precision of people's general knowledge. In addition, the recognition heuristic still predicted the lay predictions best when collective recognition, rather than individual recognition, was used. Of course, we cannot exclude that mechanisms other than those considered would have fared better. However, apart from the fact that a proportion of 90% correctly predicted forecasts is difficult to exceed, the four mechanisms we tested are plausible ones and indeed—to some extent—are useful for predicting the actual outcomes of the matches.

Regarding the applicability of the recognition heuristic, the heuristic could not, as mentioned above, be used throughout. This is because the heuristic relies on very "coarse" information (with dichotomous cues such as recognition representing the extreme high end of coarseness). As a consequence, in a reference class of N objects, the recognition cue can have a maximum applicability of 1 only with $N = 2$ objects and with increasing N, the maximum applicability quickly converges to 0.5 (Gigerenzer & Goldstein, 1996). In other words, in many domains the recognition heuristic can at most be applied in only around half of all possible comparisons. Note that this applies to every single-variable strategy that is based on a binary cue. It was not in the scope of the present study to investigate what mechanisms are used when the recognition heuristic cannot be applied (e.g., when both teams are recognized). One possibility could be lexicographic strategies such as Take The Best (Gigerenzer & Goldstein, 1996), which looks up cues in decreasing order of validity, or the fluency heuristic (Schooler & Hertwig, 2005), which uses recognition speed to discriminate between two recognized objects. Future research should examine further the role of such mechanisms in sports forecasting.

7.2. The ecological rationality of using recognition in sports forecasting

Recognition allowed participants to predict the match outcomes well above chance level (71% and 63% correct forecasts for individual and collective recognition, respectively). However, it was considerably worse than four of the five direct indicators of team strength that were tested. In other words, we were unable to replicate the results obtained by Serwe and Frings (in press), who found, in the context of tennis matches, that recognition performed similarly to or even better than rankings. It has to be noted, however, that in absolute terms, in our study, the performance of individual recognition was similar to that in Serwe and Frings (with 71% vs. 73–76% correct predictions). That is, the inferiority of recognition in our study was mainly due to the rankings reaching a higher accuracy than in the tennis domain studied by Serwe and Frings (81.3% and 84.5% vs. 68%; see Fig. 1). One speculative reason for this difference might be that the rankings of soccer teams are more robust than ATP rankings of the tennis players. In any case, in terms of predictive strength, recognition seems to be rather robust across different sports domains, whereas other indicators vary considerably.

7.3. Is partial ignorance beneficial?

We looked at two situations in which a lack of knowledge could be advantageous, leading to a less-is-more effect. In neither of them did we observe such an advantage. The rationale of the less-is-more effect is that if the recognition validity is higher than the knowledge validity and the recognition heuristic is used wherever possible, the number of instances in which the recognition heuristic can be applied decreases when more than around half of the objects are recognized. Beyond this point, the overall accuracy is determined mainly by the knowledge validity, leading to a decreasing overall accuracy—if the recognition and the knowledge validity remain relatively *constant* across different levels of n. In our study, in contrast, higher levels of n were associated with *higher* knowledge and recognition validities. Possibly, this prevented any benefit of ignorance. If this was indeed the case, our results might indicate that the condition $\alpha > \beta$ is not sufficient to produce a less-is-more effect (when the recognition heuristic is used). Examining the impact of the correlation of n with α and β on the emergence of a less-is-more effect more systematically would have required a much larger sample size than we had in our study. In general, however, the observation that n was correlated with α and β might mean that actual manifestations of a less-is-more effect will be rather rare.

We do not know exactly what information the experts had but the laypeople did not. One possibility is that the experts knew something about the teams' previous performance and their FIFA ranks, which, at least for the EURO 2004 tournament, we showed were very good indicators of the success of a team. Other studies (Serwe & Frings, in press; Andersson et al., 2003), however, have shown that expert knowledge and expert performance do not always excel, and in these situations a less-is-more effect might be more likely to emerge.

8. Conclusion

Limited knowledge does not need to stand in the way of good forecasting. In the sports context studied in this article, people with limited soccer knowledge appeared to exploit the fact that they had not heard of some teams as an indication that these teams were not as strong as those they had heard of. Using such a strategy was smart, as recognition allowed them to make correct predictions well above chance level. However, the power of recognition also has its bounds: Recognition could not predict the actual results as well as various direct indicators of team strength. Moreover, that ignorance, or lack of recognition, leads to better forecasts could not be supported.

Acknowledgments

The authors wish to thank Patric Andersson, Gerd Gigerenzer, Robert Hartsuiker, Ralph Hertwig, Robin Hogarth, John Hutchinson, Lael Schooler, and Peter Todd for constructive comments on an earlier draft of this article. Furthermore, Gregor Caregnato and Sonja Schulze are thanked for their assistance with the data collection, and Anita Todd for editing the manuscript.

References

American Gaming Association. (2004). 2004 State of the states: The AGA survey of casino entertainment. Retrieved on 09/21/04 from http://www.americangaming.org/assets/files/2004_Survey_for_Web.pdf.

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., Ekman, M., & Edman, J. (2003). Forecasting the fast and frugal way: a study of performance and information-processing strategies of experts and non-experts when predicting the World Cup 2002 in soccer. SSE/EFI Working Paper Series in Business Administration no. 2003:9. Stockholm School of Economics, Stockholm.

Ayton, P., & Önkal, D. (submitted for publication). *Effects of ignorance and information on judgmental forecasting*. Manuscript.

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.

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: Cambridge University Press.

Cantinotti, M., Ladouceur, R., & Jacques, C. (2004). Sports betting: can gamblers beat randomness? *Psychology of Addictive Behaviors, 18*, 143–147.

Condorcet, M. (1994). Essay on the application of probability analyses to decisions returned by a plurality of people. In I. McLean, F. Hewitt (Eds. & Trans.), *Condorcet: Foundations of social choice and political theory* (pp. 11–36). Brookfield, VT: Edward Elgar (Original work published 1785).

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

Forrest, D., & Simmons, R. (2000). Forecasting sports results: the behaviour and performance of football tipsters. *International Journal of Forecasting, 16*, 317–331.

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

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: on the misperception of random sequences. *Cognitive Psychology, 17*, 295–314.

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

Grofman, B., & Owen, G. (1986). *Information pooling and group decision making*. Greenwich, CT: JAI Press.

Heath, C., & Gonzalez, R. (1995). Interaction with others increases decision confidence but not decision quality: evidence against information collection views of interactive decision making. *Organizational Behavior and Human Decision Processes, 61*, 305–326.

Heit, E., Price, P., & Bower, G. (1994). A model for predicting the outcomes of basketball games. *Applied Cognitive Psychology, 8*, 621–639.

Kaplan, R. M. (1980). How do fans and oddsmakers differ in their judgments of football teams? *Personality & Social Psychology Bulletin, 6*, 287–292.

Koehler, D. J. (1996). A strength model of probability judgments for tournaments. *Organizational Behavior and Human Decision Processes, 66*, 16–21.

Ladouceur, R., Giroux, I., & Jacques, C. (1998). Winning on the horses: how much strategy and knowledge are needed? *The Journal of Psychology, 132*, 133–142.

Macy, R. (1999). Ban on college sports betting could costs state books millions. *Las Vegas Review Journal*, 4A.

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

Pachur, T., & Hertwig, R. (in press). On the psychology of the recognition heuristic: retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Phillips, J. K., Klein, G., & Sieck, W. R. (2004). Expertise in judgment and decision making: a case for training intuitive decision skills. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 295–315). Oxford, UK: Blackwell.

Reimer, T., & Katsikopoulos, K. (2004). The use of recognition in group decision-making. *Cognitive Science, 28*, 1009–1029.

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112*, 610–628.

Serwe, S., & Frings, C. (in press). Who will win Wimbledon 2003? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making*.

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

Vertinsky, P., Kanetkar, V., Vertinsky, I., & Wilson, G. (1986). Prediction of wins and losses in a series of field hockey games: a study of probability assessment quality and cognitive information-processing models of players. *Organizational Behavior and Human Decision Processes, 38*, 392–404.

Weeber, K. W. (1998). *Alltag im alten Rom*. Düsseldorf: Artemis & Winkler.