

USE OF CUES IN FORECASTING TASKS: EVIDENCE FROM SPORT PREDICTIONS

Norbert Becser

Assistant Professor, Institute of Business Economics, Corvinus University of Budapest,
Hungary

Bálint Esse

Adjunct Lecturer, Institute of Business Economics, Corvinus University of Budapest,
Hungary

Richárd Szántó

Associate Professor, Institute of Business Economics, Corvinus University of Budapest,
Hungary

Email: richard.szanto@uni-corvinus.hu

We asked Hungarian students to make soccer forecasts for a round of the Chilean Primera División, which was selected to minimise recognition heuristics, as this league is not followed by Hungarian soccer fans. Students made their forecasts based on cues about the teams. They received sets of cues of different quality based on previously composed expert evaluations of the given information. Although there were interesting findings regarding expertise and gender differences, the central question concerned how students evaluated the cues they received. Results show a pattern that we interpret as the contrast effect in information sets, which is present independent of gender or expertise.

Keywords: sports forecasting, cue value, gender differences

JEL codes: D81, D83

1. INTRODUCTION

Although sport forecasting or betting is very popular, relatively few papers deal exclusively with sport forecasting (Stekler et al. 2010). This lack of research is more surprising if one takes into account that research on judgmental forecasting has become increasingly popular over the last 25 years (Lawrence et al. 2006). Nevertheless, studying sports forecasting is a scientifically valuable task as it is possible to analyse the role of (the lack of) expertise in decision-making (Andersson 2008), individual differences and the available information cues in a specific forecasting task. Studies have tried to explore

the expert/non-expert gap in the field of sports predictions, with many finding only a narrow difference between the two groups.

We considered all three aspects in our study. We asked Hungarian students to make soccer forecasts for a round of the Chilean Primera División (i.e. the Chilean soccer league). Students made their forecasts based on cues about the teams. They received sets of cues of different quality based on previously composed expert evaluations of the given information. The central question of our study concerned how students evaluated the cues they received, and we also explored individual differences in this soccer prediction task.

This paper is organized as follows. First, Section 2 summarizes the most important findings of past studies on the performance of experts and non-experts in forecasting tasks, the gender differences in forecasting, and the role of information in sport forecasting, with special emphasis on soccer predictions. Section 3 describes the experiment that we conducted in our research. Section 4 introduces the results of our research. Finally, Section 5 presents the discussion and our conclusions that can be drawn from the results.

2. LITERATURE REVIEW

2.1. Performance of experts and non-experts in forecasting tasks

There has been a growing literature on the difference between experts and non-experts in various forecasting tasks. Andersson et al. (2005) found that the accuracy of experts and non-experts in predicting the winner of various soccer world cup final matches did not differ significantly (this finding was supported again four years later (Andersson et al. 2009), which suggest that sports experts have limited forecasting ability. Soccer tipsters (whose expertise is based on experience and intuition) are outperformed by prediction markets and betting odds according to Spann and Skiera (2009). One may claim that the reason for poor performance can be complex, arising from the lack of efficient use of available information (for e.g. newspaper tipsters not being able to adequately use obtainable information on teams' strength (Forrest and Simmons 2000), and the overall complexity of the task, etc., but Andersson et al (2009) also suggest that sports journalists, coaches or fans cannot be considered experts in sports predictions in a traditional way as forecasting is not part of their daily activity (without underestimating their previous experiences and knowledge). Stekler et al. (2010) argue that experts in this field include tipsters who write for newspapers and bookmakers who provide fixed odds. In the defense of experts, while some evidence suggests that statistical systems that predict game winners do not differ from some expert groups in terms of success rates, betting markets outperform both (Song et al. 2007), and in some observed periods subjective forecast of English odds-setters produce better predictions than some

sophisticated statistical models (Forrest et al. 2005; Andersson 2008). These suggest that the performance of some expert groups have recently improved.

The fact that laypersons often outperform (or perform at the same level as) experts can be attributed to the phenomenon that complex forecasting methods do not always provide better results than very simple ones (Goldstein – Gigerenzer 2009). Simple “fast and frugal” heuristics such as recognition, take-the-last-ranking (Raab 2012) or simple averaging may work better in certain domains than highly sophisticated expert models. Serwe and Frings (2006) found that the recognition rule (participants choose the player that they recognized) was almost as efficient as expert judgment when subjects had to predict the winners of the Wimbledon Tennis Tournament. Another similar study (Scheibehenne – Bröder 2007) revealed that recognition-based rankings are as accurate predictors of Wimbledon games as ATP rankings (which is the world ranking of professional tennis players) or the expert committee of the tournament that is responsible for official seeding in Wimbledon, while Herczog and Hertwig (2011) claim that the prediction power of the ATP rule (based on the ATP ranking) can be improved by incorporating collective recognition rankings into the forecasting models.

2.2. Gender differences in forecasting

Although soccer knowledge is part of several gender stereotypes (men are usually considered more knowledgeable about soccer), and soccer is usually considered a masculine sport in general (see for example Koivula 2001), even after an extensive search, we were unable to find past studies discussing gender differences in soccer predictions (or sport prediction exercises). It is interesting to note that when Andersson et al. (2005) compared the forecasting capabilities of experts and non-experts, the expert group (sports journalists, soccer fans and soccer coaches) included only males (52 people), hence they did not have the opportunity to investigate gender differences.

Moreover, to the best of our knowledge, very few studies discuss prediction accuracy differences between sexes in other domains (Eroglu – Knemeyer 2010), and the very limited findings usually contradict each other. Sjöberg (2009) for example experienced large gender differences in election forecasting tasks when male participants predicted the results of the Swedish Parliamentary elections with a much greater accuracy than female subjects. Dunn et al (2007) studied individual differences in affective forecasting abilities, and they found that women’s forecasting accuracy was significantly higher than men’s, from which one may argue that accuracy of affective forecasting tasks depend on emotional intelligence rather than actual forecasting skills. Although Andersson et al (2009) used gender as an independent variable in their study, they did not report any significant differences.

2.3. The role of information in sport forecasting

The use and evaluation of cues during the decision-making process have been discussed by a myriad of studies since the 1960s. For example, in an early work of Paul Slovic (1966,) subjects had to judge the intelligence of a person based on several cues such as his or her high school grade or effectiveness in English. Slovic (1966) found that cue-consistency influenced how people used information when forming a judgement. When the most valuable cues were consistent, judges used linear additive models, yet when these cues contradicted each other, they picked only one. Slovic suggests that the intrinsic validity of the cues still matters, hence a cue with low validity will be used only when it is consistent with other more valid cues. Einhorn (1971) claims that subjects with a greater amount of information use more complex combinations of linear and non-linear (for example conjunctive) models, while his experiment results did not support his original hypothesis, namely that an increasing amount of information triggers more intense use of non-linear models. Similarly, Payne (1976) asserts that task complexity does influence information processing: when few alternatives are available, people tend to use compensatory models, while in case of more alternatives subjects try to quickly eliminate some options by applying non-compensatory models in order to reduce the information processing involved in a more complex task.

The role of (additional) information has been discussed by few studies in the sports forecasting domain. The relatively low number of these studies can be explained by the fact that many of them proposed recognition heuristic as a powerful tool to predict sport event results (Serwe – Frings 2006; Scheibehenne – Bröder 2007; Pachur – Biele 2007; Goldstein – Gigerenzer 2009; Herczog – Hertwig 2011; Ayton et al. 2011). Recognition heuristics work only if no additional information is acquired by the subjects (Goldstein – Gigerenzer 2002).

Andersson et al (2009) manipulated the access of information (FIFA rankings of the soccer teams involved) in their questionnaires. Their findings suggest that participants (both experts and non-experts) foresaw match results with similar accuracy regardless of whether information was presented or not. They argue that the aforementioned task was relatively easy, and therefore information use was not really needed. Earlier studies conducted by Andersson et al. (2005) had similar findings, but some less knowledgeable participants were able to use additional information cues, i.e. world cup rankings. However, when naïve participants tried to predict exact scores of matches (a highly difficult task), they performed worse when rankings were available. Ayton et al (2011) provided their less knowledgeable participants with additional half-time score information during a different prediction task, but they did not see significant differences when subjects had to forecast in the absence of this information. Since Ayton et al. supposed that Turkish students mostly used recognition heuristics when they predicted unfamiliar British soccer match results, they also hypothesized that additional

information is only used when recognition cues concurred or conflicted with half-time scores.

Tsai et al. (2008) tested a large set of cues (past performance statistics) in their (American) football result predictions in order to see how subjects use these pieces of information. They provided the cues in blocks of six, and in each round forecasters made judgments about the results, enabling researchers to test for the accuracy and confidence level of the participants. Additional blocks of cues did not improve forecasting accuracy, but steadily raised confidence levels (which is in accordance with Ayton et al. (2011), where half-time score information did not improve prediction accuracy, but did greatly increase the confidence of the participants). According to Tsai et al. (2008) cue quality matters, as when the best cues were shown first, participants exhibited relatively high confidence levels that did not increase much throughout the trials, while when poor cues were started with, confidence levels were lower at the beginning, and rose faster.

Findings of studies mentioned above contradict the fundamental assumptions of most economic discussions on information, namely, that more information produces better judgment and/or decision- making (Loewenstein et al. 2006). This assumption has been challenged several times (see for example the works of Jacoby et al. (1974), Keller and Staelin (1987), or Goldstein and Gigerenzer (2002)), but it seems to hold in standard economic studies.

2.4. Relevant cues in soccer prediction

What counts as relevant information in soccer prediction tasks? Some studies reveal that even shirt colour in various sport contests can influence results (most of the time due to visibility biases). Yet, Kocher and Sutter (2008) argue that in soccer, or more generally in team sports, jersey colour itself does not affect teams' performance. Information on a team's past performance seems to be important in forecasting (Pachur – Biele 2007), but forecasters do not necessarily exploit this knowledge. In a different domain, Song et al. (2007) investigated two seasons of the National Football League (NFL) and found that in the second half of each season the forecasting accuracy of football experts (former players, sports journalist) declined. This finding is unexpected as in the second half of the season one can assume that more information is available for tipsters to use. Nonetheless, in a recent soccer prediction study about FIFA World Cup and the German Soccer League results (Raab 2012), several cues from the previous season such as final ranking, winning percentage, home advantage, and goals scored were presented to participants, who then used only the rank in the final standings. The findings of Goddard (2006) suggest that newly introduced managers will not necessarily lead to better team performance, while the home field advantage is not just a myth, as on average teams playing at home win more matches than away. Moreover, distance also matters: the greater the distance a team has to travel the smaller the chances are for an away-game win (at least in the English

soccer league). Last, Boulrier and Stekler (1999) argue that sports seedings (and rankings in general) are useful predictors of future performance in the field of basketball and tennis (the latter finding, i.e. that ATP and WTA rankings are good predictors for Grand Slam tennis outcomes was also supported by del Corral and Prieto-Rodríguez (2010)).

In consequence, one may conclude that providing information to participants does not necessarily result in better forecasting performance. The authors of the aforementioned studies did not investigate how useful participants found the cues, or in other words, what the perceived value of information was. Yet, it can be assumed that participants may evaluate the available cues differently from the perspective of their predictive power. According to the dominant view, standard economic information is valuable only to the extent that it is instrumental for decision- making (Eliaz – Schotter 2010). However, forecasters may underestimate or overestimate the usefulness of the given cues depending of the presentation format or the number of cues.

3. EXPERIMENT

We tested how participants perceive the usefulness of provided cues in forecasting the results of nine matches of a round in the Chilean Primera División.

3.1. Method

Although there are nine matches, this can be considered as a one-round judgemental forecasting task. Materials provided in the study were questionnaires. These included various cues about the teams, the list of matches in the actual round, scaling questions on the perceived value of information, questions about the country in general, and questions about expertise in sports betting.

All questionnaires were identical, except the cues provided: there were four different sets. These were the following:

- 0i: no information;
- 5i: five irrelevant cues (age of the goalkeeper, citizenship of the coach, jersey colour, inhabitants of the city, yellow cards collected so far);
- 5r: five relevant cues (total value of players in EUR, number of national players in the team, current winning or losing streaks, current position in league, final position in league last season);
- mixed: the previous two groups together, five relevant and five irrelevant cues.¹

¹ We initially made two mixed sets: one, where the irrelevant cues were followed by relevant ones, and one in reverse order. Our question was whether the order of exposure to information changes anything. It did not, as there was no significant difference between these two sets. In consequence, we merged the two sets, as there was no value added in presenting them separately.

Information on all eighteen teams were presented in a table, except the group forecasting the results without information. The sources of information were the webpages of the teams, their pages on social media sites and international soccer web-databases.

All participants, including those with no information, evaluated the usefulness of the cues listed. The subsample without information evaluated how useful all the information would have been. After these, five questions about Chile followed. These were just proxies of the knowledge of the environment, to exclude the possible effect of recognition heuristic (Raab 2012). Chile had been chosen because it was rather improbable that Hungarian students would know any of the team names, minimising the effect of recognition heuristics.

Data on perceived soccer expertise, experience in soccer forecasting, expected number of correct forecasts, age and gender were also collected from the participants of this self-reported questionnaire. Questionnaire data was processed with SPSS software. Due to data distribution, the Kruskal-Wallis test was used to check the presence of significant differences between groups, and Mann-Whitney U test to identify these exactly.

3.2. Hypotheses

The following three hypotheses were formulated and tested in the study:

H1: The perceived usefulness of a cue varies depending on the set in which it is presented.

This was the main hypothesis of the study. Different information cue sets were prepared with different compositions to examine how the perceived value of information changes. We expected that the value of information would depend on the amount of information in the set, as well as on the perceived value of other cues.

H2: Different cue sets lead to different results: a better result can be achieved when information is provided.

It was assumed, that with (good quality) information the results of more matches would be predicted correctly.

H3: Experts reach better results.

In the study, participants rate their own expertise on a 5-point scale. In this hypothesis the

expectation is expressed that those rating themselves higher on this scale reach better results: more hits in match results.

3.3. The sample

As part of the study, we worked with an expert sample. Before the experiment 37 cues were collected that could possibly be useful in predicting the results of soccer matches. These were then ranked by 27 experts according to the perceived forecasting power of the cues in soccer. The expert group consisted of sports journalists, coaches, former players and professors of sports management. Based on the answers from this sample we labelled cues as *relevant* or *irrelevant*, which were the two ends of the rankings.

Participants were university students, without any financial or other incentives, except that the best bettors in all categories received a reward. The study was conducted at a management and business administration program of a university in Budapest, Hungary. There were 203 participants in the study, 90 of them female. The average age in the sample was 21.37, with a standard deviation of 1.6.

4. RESULTS

The chance of guessing a match result right without knowing any information is 1/3, as every match has 3 outcomes: home win, draw and away win. We could consider then our participants' guesses as better than random guesses, if their hits were more than 3 on average, which was the case. On the long term, the probabilities of the three outcomes differ: the least probable in soccer are draws, and home wins have a little higher probabilities than away wins. In this one special set of matches, there were 4 home wins (1 surprising, according to odds) and 5 away wins (1 surprising) from 9 matches. Although these are true, this does not affect the fact that the expected number of hits is without information was 3. A few participants used their expertise to guess all matches as home wins, and many participants were rather avoiding draws. These two tendencies made it possible that the average is higher than 3.

4.1. Gender and expertise differences

Male and female participants showed significant difference in their self-perceptions regarding their own expertise in soccer and betting. As presented in Table 1, average female expertise is much lower compared to men ($U=1819.50$, $p=0.000$), and in the case of expected number of hits and misses we can observe a similar difference ($U=2818.00$, $p=0.000$). However, the average number of actual hits show only differences that are close to significant ($U=4312.00$, $p=0.064$). We analysed these relationships also across questionnaire types, where the same relationships appear (except in case of the 5r type,

where there is no difference in the expected number of hits).

Table 1. Gender differences in perceived expertise, hits and expected number of hits

	Male	Female
Average number of hits	3.69 sd=1.01	3.99 sd=1.16
Perceived expertise (scale 1-5)	3.18** sd=1.32	1.62** sd=0.91
Expected number of hits	5.28** sd=1.55	4.06** sd=1.69

Source: authors

We can conclude that in this one-round forecasting task, there was a significant difference between male and female participants in the perception of their own expertise and also in the expectations about their hits. Female participants were less confident than males. Despite this, there is no significant difference between their actual results in this round. We have to add that in the context of Hungary's national sport culture, males are more inclined to watch soccer, to bet, and men's soccer is far more popular than women's.

4.2. The role of own perceived expertise

If we accept that even in a one-round task experts should reach more hits, then we would expect a clear positive relationship between these two variables. When hits and expertise were compared, the averages showed a seemingly interesting U-shape: those indicating low and high expertise have more hits than the middle. We could speculate about a group underestimating their own expertise (experts knowing – as part of their expertise – that in a one-round task they are not able to show their expertise), but from this one shot task this is not in evidence, but can be an interesting future research direction.

Table 2. Levels of own perceived expertise and the number of hits

Expertise	Average number of hits	1	2	3	4	5
1	4.20		U=1105.00 p=0.007	U=1091.50, p=0.002	U=580.00, p=0.001	-
2	3.55			-	-	-

3	3.60				-	-
4	3.46					U=157.50, p=0.061
5	3.95					

Source: authors

It is interesting at first sight how the group indicating the least expertise reached the highest average number of hits. After testing the significance of differences we can conclude that there are significant differences present (Chi-Sq=18.11, $p=0.001$). From pairwise comparison of expertise levels it appears that the average number of hits from the lowest level group in expertise differs significantly from all other groups, except from the highest level, and the highest level group's average differs significantly only from expertise level 4. We must reject hypothesis 3: in this one round task subjects considering themselves experts did not reach better results.

When we compare expertise with the expected number of hits, our participants are quite confident. The higher the expertise, the higher the expected number of hits (from 4.16 to 5.26, Chi-Square=18.108, $p=0.001$). Testing the actual differences with neighbouring values shows no significant difference, while differences are significant between two step leaps (expertise 1 and 3, 2 and 4, and so on).

4.3. The various cue sets and the value of information

As already presented, different cue sets were provided to participants that could be used in making forecasts. The first question concerned whether it pays off at all to use this information: do those receiving relevant information guess more results correctly in one round? Result are shown in Table 3.

There are significant differences between groups (Chi sq.=11.847, $p=0.008$), which means that cue sets mattered to achieve a good hit rate even in this one-round task. As presented in Table 3, there is no difference between the results arrived at with zero information and when participants received five pieces of relevant or irrelevant pieces of information. With ten pieces of information the number of hits was significantly higher. Also, the results with any other kind of information were better than with five irrelevant cues. Our results may be slightly ambiguous, but we accept hypothesis 2: different cue sets resulted in different results. However, we must emphasise that the results do not show the exact pattern expected: all information is better than irrelevant information, and there is an almost significant difference between relevant and irrelevant cues, but at the same time, only the mixed cue set helped to reach significantly better results than no information.

Table 3. Cue sets, the average number of hits and pairwise significant differences

Cue set	hits	0i	5i	5r	mixed
0 info	3.49		-	-	(U=1170.50 p=.009)
5 irrelevant	3.55			(U=660.50, p=.056)	(U=1105.50, p=.005)
5 relevant	4.00				-
mixed	4.10				

Source: authors

Careful attention was paid to how useful the cues were found by participants. When we checked their evaluations, the following patterns emerged:

- subjects who received only irrelevant information evaluated these higher on a five-point scale compared to those who received an additional five relevant cues. There is one exception: it seems that jersey colour is so evidently irrelevant that it resulted in the same value in all cue sets;
- subjects who worked with five relevant cues attached consequently lower values to these cues than participants working with the mixed pool: adding five irrelevant cues made them value the relevant cues more;
- participants with no information, who could only imagine what the extra information consisted of, showed almost the same curve as those having the mixed set.

This effect can be interpreted as the *contrast effect* (Plous 1993) in information valuation: the perceived value of information changes when it is presented together with other information of significantly different quality. Exact values and the significance of these differences is presented in Table 4. When we look for significant differences, we find that except for two pieces of information (jersey colour and W/L streak – $p=0.060$), there are differences in the mean compared to the mixed set across eight cues. Thus, the first hypothesis can be accepted: the value of the same information piece varies depending on which set it is presented in.

Table 4. The cue sets and the value of information

	0i	5r	5i	Mixed
Age of the	2.12		2.95	1.76

goalkeeper	U=500.50, p=0.003 to 5i			U=684.50, p=0.000 to 5i U=1360.50, p=0.125 to 0i
Citizenship of the coach	2.37 U=752.50, p=0.638 to 5i		2.23	1.39 U=760.00, p=0.000 to 5i U=783.50, p=0.000 to 0i
Inhabitants of the city	2.20 U=466.50, p=0.001 to 5i		3.21	1.58 U=558.00, p=0.000 to 5i U=1111.00, p=0.002 to 0i
Jersey colour	1.27 U=765.50, p=0.598 to 5i		1.21	1.13 U=1461.00, p=0.367 to 5i U=1467.50, p=0.119 to 0i
Yellow cards collected so far	2.56 U=611.50, p=0.059 to 5i		3.08	1.99 U=728.00, p=0.000 to 5i U=1030.00, p=0.001 to 0i
Total value of players (in EUR)	4.32 U=157.50, p=0.000 to 5r	2.42		3.28 U=964.50, p=0.000 to 5r U=704.00, p=0.000 to 0i
Winning or losing streaks	4.27 U=808.00, p=0.475 to 5r	4.09		4.47 U=1366.00, p=0.060 to 5r U=1429.50, p=0.283 to 0i
Number of national players in the team	4.29 U=256.00, p=0.000 to 5r	2.93		3.45 U=1219.00, p=0.009 to 5r U=857.50, p=0.000 to 0i
Current position in the league	4.51 U=624.50, p=0.020 to 5r	3.93		4.62 U=1096.50, p=0.001 to 5r U=1508.50, p=0.473 to 0i
Position in league at the end of previous season	4.00 U=620.00, p=0.012 to 5r	3.44		4.03 U=1135.00, p=0.002 to 5r U=1503.50, p=0.496 to 0i

Source: authors.

After finding the contrast effect in data, we were curious about how this could be explained. Men and women do not show any difference in this pattern. When we analyse the results of cue values, or the average value of all cues by one participant, the pattern is the same across the sexes. When comparing expertise and the value of cues, the results

are the same. We would expect experts to be less prone to evaluating cues with this pattern, but were surprised to see it did not happen. We considered those indicating higher expertise than 3 as “experts”, and less than 3 as “non-experts”. No matter whether we defined it this way, or analysed the whole expertise scale, the pattern remained the same. It seems that experts are also susceptible to the contrast effect. Participants guessed their correct guesses in the end. We thought this might relate somehow to cue valuation as well: those expecting a high number of hits would value their cues more highly (this way their expectation and the perceived values would be consonant), but this did not happen.

To conclude, we were unable to explain the above presented pattern with differences in gender, expertise or confidence. It seems plausible that the contrast effect is a general phenomenon, and expertise cannot prevent it.

5. DISCUSSION AND CONCLUSIONS

An intensive literature review revealed that gender differences in forecasting tasks have hardly been investigated in the past, and that the relatively limited findings are rather subtle and contradictory. This is fairly surprising given the findings of this study, where expected accuracy and confidence of males and females significantly differed in most cases. Hence, further research avenues should include studies on gender differences using various domains of forecasting. As soccer is typically considered a male-centered domain where differences may be expected, other domains where gender stereotypes do not exist should also be examined.

Participants were asked to rate their expertise in soccer. The results show a U shape: the best results were achieved by participants who rated themselves as laypeople in this field. Although this was a one-round task, an interesting question arises: how do people rate their own expertise? Is the good hit rate the result of expert predictions, whereby “humble experts” know how small their chances are, but still use the available information well to achieve a good result? There is thus a need to examine within other domains and tasks how people rate their own expertise. It has been proven (see for example Jakobsson et al. 2013 or Barber – Odean 2001) that overconfidence is task-dependent, and the difference between male and female participants disappear as soon as they face neutral tasks, not overtly masculine ones, such as soccer prediction. But what about self-perceptions and evaluations of expertise? Are these stable, general views of one’s capabilities? Or are they task-dependent? In our example: does knowing that the chances of success are poor (which experts know) have an effect on how a participant rates her expertise? Are there confident experts and humble experts, or experts hiding their belief about their own expertise because of fears of failure? Research in these areas would be able to uncover the initial self-confidence in tasks where expert knowledge is needed.

In our work, different cue sets resulted in different numbers of hits, but expertise did not. The most interesting finding of the study is that the perceived value of the provided cues depends on the pool in which these are presented. Relevant information was found to be more valuable when presented with the same amount of irrelevant cues, and vice versa. This pattern of distortion in information evaluation cannot be explained by expertise, gender, or confidence differences.

A very compelling finding is in relation to subjects who with zero information assigned almost the same values to cues as those who received all available information. This is important because these people did not see the actual information that was available, such that they could only evaluate their hypothetical value. Where does this „knowledge” disappear when the contrast effect arises? As these participants only saw what cues they *could* get, it was not the information (its variance) that counted, but in knowing the cue labels. As soon as they received only relevant, or only irrelevant information, the contrast effect appeared. An outstanding question concerns whether this would happen if these participants did not receive five (ir)relevant pieces of information, but just knew their labels/types.

We identified this biased evaluation of cues with the contrast effect. This is similar to anchoring bias, but while in the case of anchoring the initial stimulus is on the same attribute as the answer of the respondent, here the perceived value of one cue distorts the value of another. This is also similar to the decoy effect, where with the latter our preference between two or more options is biased by the introduction of a third option, with the contrast effect no inferences are made (or at least were not asked for). How may this contrast effect work in other domains, (e.g. performance appraisal), particularly with regard to differential effects on groups? In our study, expertise did not matter. What may count, then? Is this a systematic error, or are there contextual or task-relevant factors that can account for this finding? It is relevant to note that there are cues which cannot confuse participants (in our case the jersey colour, receiving poor values in all sets), implying that the qualities of misleading information also need to be investigated and accounted for.

In a more realistic setting, where participants make predictions every week (i.e. in each round of a soccer championship), a learning curve may play a significant role in predictions: additional information about standings, players, results etc. will influence the subjects' predictions. In reality, the evaluation of information cues is more dynamic, as people regularly reassess the value of the information they possess. It is possible that in the long run the magnitude of the contrast effect could diminish or disappear, as participants, building on their experiences from previous rounds, adjust their evaluations, and irrelevant items in their information pool do not greatly influence the perceived value of cues. Nevertheless, our findings have some practical implications for the betting industry. The way how betting sites, sports magazines, tipsters, bookmakers etc. provide cues for bettors can influence the bettors' behavior, therefore these information sources

may have a greater role and responsibility than one might think.

One must note that the level of uncertainty of the outcome of soccer matches is highly dependent on various factors, for example on the value of the teams playing – top teams' (with star players) winning chance is better when playing against an average team. Several initiatives could increase the uncertainty level of the matches such as using a salary cap, or applying special rules (for example basic financial and non-financial requirements for being eligible to enter the league). A salary cap sets the maximum amount of payment that a star player can receive. Using the salary cap – as it is used for example in the Major League Soccer in the US – could lead to increase the fairness of competition. Top players could be paid by not only larger, wealthy clubs, but also other clubs could compete in the transfer market. That could lead to a more balanced composition of leagues and of course, to more exciting matches with increased uncertainty.

REFERENCES

- Andersson, P. (2008): Expert predictions on football: A survey of the literature and an empirical inquiry into tippsters' and odds-setters' ability to predict the world cup. In: Andersson, P. – Ayton, P. – Schmidt, C.: *Myths and facts about football: The economics and psychology of the world's greatest sport*. Cambridge Scholar Publishing.
- Andersson, P. – Edman, J. – Ekman, M. (2005): Predicting the World Cup 2002: Performance and confidence of experts and non-experts. *International Journal of Forecasting* 21(3): 565–576.
- Andersson, P. – Memmert, D. – Popowitz, E. (2009): Forecasting outcomes of the World Cup 2006 in football: Performance and confidence of bettors and laypeople. *Psychology of Sport and Exercise* 10(1): 116–123.
- Ayton, P. – Önköl, D. – McReynolds, L. (2011): Effects of ignorance and information on judgments and decisions. *Judgment and Decision Making* 6(5): 381–391.
- Boulter, B.L. – Stekler, H.O. (1998): Are sports seedings good predictors?: An evaluation. *International Journal of Forecasting* 15(1): 83–91.
- del Corral, J. – Prieto-Rodríguez, J. (2010): Are differences in ranks good predictors for Grand Slam tennis matches? *International Journal of Forecasting* 26(3): 551–563.
- Dunn, E.W. – Brackett, M.A. – Ashton-James, C. – Schneiderman, E. – Salovey, P. (2007): On emotionally intelligent time travel: Individual differences in affective forecasting ability. *Personality and Social Psychology Bulletin* 33(1): 85–93.
- Einhorn, H.J. (1971): Use of nonlinear, noncompensatory models as a function of task and amount of information. *Organizational Behavior and Human Performance* 6: 1–27.

- Eliaz, K. – Schotter, A. (2010): Paying for confidence: An experimental study of the demand for non-instrumental information. *Games and Economic Behavior* 70(2): 304–324.
- Eroglu, C. – Knemeyer, A.M. (2010): Exploring the potential effects of forecaster motivational orientation and gender on judgmental adjustments of statistical forecasts. *Journal of Business Logistics* 31(1): 179–195
- Forrest, D. – Simmons, R. (2000): Forecasting sports results: the behaviour and performance of football tipsters. *International Journal of Forecasting* 16(3): 317–331.
- Forrest, D. – Goddard, J. – Simmons, R. (2005): Odds-setters as forecasters: The case of English football. *International Journal of Forecasting* 21(3): 551–564.
- Goddard, J. (2006): Who wins football. *Significance* 3(1): 16–19.
- Goldstein, D. G. – Gigerenzer, G. (2002): Models of Ecological Rationality: The Recognition Heuristic. *Psychological Review* 109(1): 75–90.
- Goldstein, D. G. – Gigerenzer, G. (2009): Fast and frugal forecasting. *International Journal of Forecasting* 25(4): 760–772.
- Herczog, S.M. – Hertwig, R. (2011): The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. *Judgment and Decision Making* 6(1): 58–72.
- Jacoby, J. – Speller, D.E. – Berning, C.K. (1974): Brand Choice Behavior as a Function of Information Load. *Journal of Consumer Research* 11(1): 33–42.
- Keller, K.L. – Staelin, R. (1987): Effects of Quality and Quantity of Information on Decision Effectiveness. *Journal of Consumer Research* 14(2): 200–213.
- Kocher, M.G. – Sutter, M. (2008): Shirt colour and team performance in football. In: Andersson, P. – Ayton, P. – Schmidt: Myths and facts about football: The economics and psychology of the world's greatest sport. Newcastle upon Tyne: Cambridge Scholar Publishing.
- Koivula, N. (2001): Perceived characteristics of sports categorized as gender-neutral, feminine and masculine. *Journal of Sport Behavior* 24(4): 377–393.
- Lawrence, M. – Goodwin, P. – O'Connor, M. – Önköl, D. (2006): Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting* 22(3): 493–518.
- Loewenstein, G. – Moore, D.A. – Weber, R.A. (2006): Misperceiving the value of information in predicting the performance of others. *Experimental Economics* 9(3): 281–295.
- Pachur, T. – Biele, G. (2007): Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events. *Acta Psychologica* 125(1): 99–116.
- Payne, J.W. (1976): Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis. *Organizational Behavior and Human Performance* 16(2): 366–387.
- Plous, S. (1993): *The psychology of judgment and decision making*. McGraw-Hill.

- Raab, M. (2012): Simple heuristics in sports. *International Review of Sport and Exercise Psychology* 5(2): 104–120.
- Scheibehenne, B. – Bröder, A. (2007): Predicting Wimbledon 2005 tennis results by mere player name recognition? *International Journal of Forecasting* 23(3): 415–426.
- Serwe, S. – Frings, C. (2006): Who will win Wimbledon? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making* 19(4): 321–332.
- Sjöberg, L. (2009): Are all crowds equally wise? A comparison of political election forecasts by experts and the public. *Journal of Forecasting* 28(1): 1–18.
- Slovic, P. (1966): Cue-consistency and cue-utilization in judgment. *The American Journal of Psychology* 79(3): 427–434.
- Song, C. – Boulier, B. L. – Stekler, H. O. (2007): The comparative accuracy of judgmental and model forecasts of American football games. *International Journal of Forecasting* 23(3): 405–413.
- Spann, M. – Skiera, B. (2009): Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters. *Journal of Forecasting* 28(1): 55–72.
- Stekler, H. O. – Sendor, D. – Verlander, R. (2010): Issues in sports forecasting. *International Journal of Forecasting* 26(3): 606–621.
- Tsai, C. I. – Klayman, J. – Hastie, R. (2008): Effects of amount of information on judgment accuracy and confidence. *Organizational Behavior and Human Decision Processes* 107(2): 97–105.