In press. *Organizational Behavior and Human Decision Processes*

The illusion of knowledge: When more information reduces accuracy and increases confidence

Crystal C. Hall, Lynn Ariss, and Alexander Todorov

Princeton University

Address correspondence to Alexander Todorov, Department of Psychology, Green Hall, Princeton University, Princeton, NJ 08544. E-mail: atodorov@princeton.edu. Phone 609-258-7463. Fax: 609-258-1113.

Abstract

Intuition suggests that having more information can increase prediction accuracy of uncertain outcomes. In four experiments, we show that more knowledge can decrease accuracy and simultaneously increase prediction confidence. Participants were asked to predict basketball games sampled from a National Basketball Association season. All participants were provided with statistics (win record, halftime score), while half were additionally given the team names. Knowledge of names increased the confidence of basketball fans consistent with their belief that this knowledge improved their predictions. Contrary to this belief, it decreased the participants' accuracy by reducing their reliance on statistical cues. One of the factors contributing to this underweighting of statistical cues was a bias to bet on more familiar teams against the statistical odds. Finally, in a real betting experiment, fans earned less money if they knew the team names while persisting in their belief that this knowledge improved their predictions.

People generally believe that the more information they have the better their decisions will be (Schwartz, 2004), and will even pursue information that is inconsequential to their final decisions (Bastardi & Shafir, 1998). In this paper, we argue that more knowledge can reduce the accuracy of prediction of uncertain outcomes and simultaneously increase confidence in prediction. We focus on the prediction of sports outcomes for two reasons. First, these are uncertain events about which people have general knowledge. Second, it is possible to randomly sample these events. As critics of the heuristics-and-biases approach have pointed out, demonstrations that a bias leads to suboptimal performance require random sampling of events from the domain of prediction (Gigerenzer, Hoffrage, & Kleinbolting, 1991). Otherwise, suboptimal performance can be attributed to a biased sample of events.

The question of accuracy of human judgments is critical for the evaluation of the optimality of these judgments. Most research on biases has compared human judgments to normative models (Todorov, 1997). Such models satisfy coherence constraints and systematic deviations from their predictions reveal internal inconsistencies or biases in judgments. However, the question of internal consistency is conceptually independent of the question of accuracy (Hammond, 1996). The accuracy of judgments is assessed against an external criterion – the real outcome of the predicted event – and one can find task domains where inconsistent judgments are as accurate as consistent judgments derived from rational models (Gigerenzer & Goldstein, 1996; Gigerenzer, Todd, and the ABC Research Group, 1999). That is, biased judgments are not necessarily less accurate than unbiased judgments.

As noted above, a fair test of the accuracy of judgments requires random sampling of events in naturalistic environments. In all experiments, we randomly sampled National Basketball Association (NBA) games. For each game, participants were provided with the same statistical information: the number of games the teams won during the season and the score at halftime. The critical manipulation was that participants were either given the team names (e.g., NY Knicks vs. NJ Nets) or not given these names (Team A vs. Team B). Intuition suggests that names should help predictions to the extent that participants have specific knowledge about the teams, because names can grant access to information that may not be available in the statistical information. However, this can be precisely the reason why knowledge of team names can reduce the accuracy of prediction. For example, after the New York Knicks reached the NBA finals in 1999 against unlikely odds, the third author started believing that they are a team that can always come back. If a person with this belief is asked to predict the outcome of a game in which the Knicks are losing at halftime, he or she is likely to predict that the Knicks would win despite evidence that trailing early in the game is highly predictive of losing the game (Cooper, DeNeve, & Mosteller, 1992).

This anecdotal failure of prediction illustrates a general case in which people can treat the predicted event as unique and not amenable to statistical generalizations (Einhorn, 1986; Kahneman & Lovallo, 1993). The lessons from studies comparing clinical and actuarial predictions are instructive. These studies have shown that simple statistical models do better in predicting patients' outcome than expert judgments of clinicians (Dawes, Faust, & Meehl, 1989). In fact, models based on the judgments of experts do better than the judgments of the experts themselves (Goldberg, 1970; Dawes, 1971). One of the reasons for the suboptimal performance of clinicians is that they treat each case as unique and import knowledge irrelevant to the prediction of the outcome. In a similar fashion, knowledge of team names can bias the predictions of the game outcome. Even when provided with statistical information, decision makers may not be completely consistent in their predictions. However, when provided with the team names, they could be even less consistent. The key question is whether this inconsistency leads to less accurate predictions when games are randomly sampled.

More-is-less versus less-is-more in predictions of uncertain outcomes

The studies on clinical and actuarial predictions, as well as studies on choice among multiple options (Iyengar & Lepper, 2000; Schwartz, 2004), have emphasized the negative effects of having knowledge on decisions – the more-is-less effect. In contrast, a number of recent studies have emphasized the positive effects of *not* having knowledge on decisions (Gigerenzer et al., 1999; Goldstein & Gigerenzer 1999, 2002; Hertwig & Todd, 2003; Schooler & Hertwig, 2005). Specifically, lack of knowledge about an event can have a diagnostic value for the prediction and can be exploited by efficient heuristics that are optimal given the processing constraints on the cognitive system.

A perfect example of this tradition of research is the recognition heuristic (Goldstein and Gigerenzer, 1999, 2002). According to this heuristic, people exploit their lack of knowledge to arrive at an accurate judgment – the less-is-more effect. For example, in decisions comparing two outcomes, if one of the outcomes is recognized and the other is not, one should assume that the recognized entity has the higher value (but for an alternative view see Oppenheimer, 2003). In fact, consistent with the predictions of this heuristic, American participants made more accurate decisions about the relative size of German than of American cities (Goldstein and Gigerenzer, 2002). Decisions about pairs in which one of the cities is not recognized (more frequently the case for German than American cities in this case) are easier (predict the recognized city) than decisions in which both cities are recognized.

While the more-is-less and the less-is-more research traditions differ with respect to the implications of the behavioral findings for models of decision-making, they clearly show that, contrary to intuition, more knowledge does not necessarily improve decisions. In terms of descriptive models of decision-making, one question that is critical for the evaluation of these implications is whether decision-makers would like to have this additional knowledge. Although the less-is-more research provides valuable insights about how the cognitive system can exploit the informational structure of the decision environment, to the extent that people would like to have additional information and would use it in non-optimal way in their decisions, charitable interpretations of the less-is-more effect in terms of the optimality of human judgments are questionable.

The illusion of knowledge

More information often increases confidence in judgments even when the accuracy of judgments is not affected (Arkes, Dawes, & Christensen, 1986; Gill, Swann, & Silvera, 1998; Oskamp, 1965; Stewart, Heideman, Moninger, & Reagan-Cirincione, 1992). Heath and Tversky (1991) have shown that people prefer to bet on events about which they have some expertise than on random chance events or on events they feel ignorant about. Team names, by cueing specific knowledge, can increase the sense of expertise for basketball fans which, in turn, can increase their confidence.

Moreover, we argue that cueing specific knowledge of the teams would render the statistical information less salient (Nisbett, Zukier, & Lemley, 1981; Zukier, 1982)[1] and, ultimately, decrease the accuracy of predictions. Our work extends previous work by showing how additional information affects predictions in non-optimal ways in a well-defined domain of prediction. Not only people knowledgeable about the domain become more confident, but they also become less accurate. The additional information introduces systematic biases that reduce the accuracy of predictions. This constellation of factors – worse but overconfident decisions supported by explicit beliefs that additional knowledge improves decisions – can produce powerful and persisting illusions of knowledge. We address each of these factors – accuracy, confidence, and beliefs – in our experiments.

Overview of experiments

In Experiment 1, we demonstrate that knowledge of team names increases the confidence in predictions of basketball fans. In Experiment 2, we demonstrate that this knowledge reduces the accuracy of predictions of both fans and non-fans despite the fans' beliefs that this knowledge improves their predictions. In Experiment 3, we show that the familiarity of the teams systematically biases predictions. Finally, Experiment 4 demonstrates that, when betting on their own predictions (with their own potential winnings at stake), basketball fans earn less when presented with the team names in addition to the statistical information, even in a setting where feedback about judgments is presented.

After we report the basic phenomena in the experiments, we present analyses across all experiments addressing the underlying judgmental mechanisms. In the case of confidence, we argue that knowledge of team names cues specific knowledge about the teams and that this knowledge inflates confidence in predictions. This claim is supported by a) the finding that knowledge of team names increases confidence of only those participants knowledgeable about basketball and b) correlational and regression analysis demonstrating that frequency of watching basketball on television is positively correlated with confidence in predictions, but only when the team names are provided.

In the case of accuracy, we make two broad claims about the observed more-is-less effect. First, we argue that knowledge of team names reduces the weight of statistical cues on predictions and that this reduces the overall accuracy of both fans and non-fans. Because each participant predicted a number of games, we could compute the influence of statistical information on their predictions. We show that this influence is significantly reduced when team names are provided. We also show that the decreased reliance on statistical information in the names condition mediates the effect of team names on accuracy. The second claim is that knowledge of team names does not simply introduce random noise in predictions but introduces systematic biases. We address one such bias, namely predicting that more familiar teams are more likely to win even when the statistical odds clearly favored the less familiar teams.

Validity of statistical cues

To show that the statistical cues (win record and halftime score) are valid predictors of the outcomes of the games, we analyzed all games for the 1996/97 NBA season ($n = 1186$; information for 4 games was missing; see Todorov, 2002). The games that were used in the subsequent experiments were from later NBA seasons and, thus, the estimates of cue validities are unbiased with respect to the games predicted by our participants.

The point-biserial correlation between the outcomes of the games and the difference in wins was .52, $p < .001$. The team with better record won in 72.8% of the games. Similarly, the correlation between the outcomes and the score at halftime was .52, $p < .001$. The team leading at the halftime won in 70.9% of the games. Thus, both statistics were valid predictors of the final outcome. The correlation between the difference in wins and halftime score was .38, $p < .001$, reflecting the fact that teams with better season records were more likely to lead at the halftime.

With two predictors, there are two types of games: games with consistent predictors and games with inconsistent predictors. It is important to consider these two types of games separately because they require different prediction strategies. For 768 out of the 1,187 games, the difference in wins and the halftime score were consistent. These were the games where the team with better record was leading at halftime. For the remaining 419 games, the difference in wins and the halftime score were inconsistent. Whereas in the case of consistent predictors participants do not need to take into account both statistical cues, in the case of inconsistent predictors they need to take both cues into account. When the predictors were consistent, they predicted accurately 86.5% of the games. This number can be interpreted as indicating the ceiling of the accuracy that one can achieve in predicting the games. When the predictors were inconsistent, the accuracy rate for both the difference in wins (49.6%) and halftime score (47.0%) was not significantly different from chance (50%).

Thus, in the latter situation, one needs to incorporate information from both predictors to achieve accuracy better than chance. We used a binary logistic regression to predict the outcomes of the games[2]. This regression model predicted accurately 66.6% of the inconsistent games, a rate significantly better than chance, χ^2 (1) = 46.11, $p < .001$. This accuracy rate is inflated because the estimation and the predictions were done on the same data set. To obtain a less biased accuracy rate, we used a double cross validation procedure. The games were randomly divided into two samples of 593 and 594 games respectively. One of the samples was used for estimation and the other was used for prediction and vise versa. The accuracy rate, computed as the average of the two accuracy rates from the double cross-validation, was 65.4%. Thus, it was possible to predict games with inconsistent predictors better than chance.

In Experiment 1, we sampled only games with inconsistent predictors. In Experiments 2 and 4, we sampled randomly from the pool of all games. In Experiment 3, we used a non-randomly selected set of games to address one determinant of the bias introduced by knowledge of team names, as well as a randomly selected set of games.

Experiment 1

Participants predicted games with inconsistent predictors – win record favoring one of the teams and halftime score favoring the other team. We started with a selection of these games for two reasons. First, when participants are presented only with statistical cues, games with consistent predictors introduce little uncertainty and participants' predictions would have been very easy. Second, games with inconsistent predictors introduce a lot of uncertainty, as shown above, and knowledge of the team names, in principle, could improve predictions. Participants who are familiar with the NBA may use the names as an additional source of information to resolve the inconsistency of the predictors. Whether or not this knowledge improves the prediction of these participants, it should increase their confidence.

In principle, it is possible that participants knowledgeable about the NBA season can remember specific games (and their outcomes) when provided with the names of the teams. However, given that there are over 1000 games in a single season, it is unlikely that these participants would have remembered many, if any, games. Moreover, memory for particular games works against our predictions that the additional knowledge of the names would decrease the accuracy of predictions.

Method

Participants. Eighty undergraduate students from Princeton University volunteered for the study, conducted in a paid questionnaire session. Participants were randomly assigned to one of two experimental conditions.

Game selection. Each NBA team plays 82 games in the regular season. For each of the 29 NBA teams, we randomly sampled 4 games from the season completed previous to the study. For each game, we recorded the number of games the teams won in the season, the score at halftime, and the final outcome. In this experiment, we selected only games in which the two statistics – team records and score at half – were inconsistent. For example, if the team record favored one of the teams (Team A won 49 games and Team B won 41 games), the score at halftime favored the other team (Team B is ahead 4 points at halftime). Using this criterion and sampling without replacement resulted in a total sample of 29 games from the initial random sample of 116 games.

Procedure. All participants were informed that each NBA team plays 82 games in the regular season and that the games to be predicted were randomly sampled from the previous season. Participants were not informed of the details of the sampling procedure in order to avoid influencing their decision strategy. It should be noted that in the remaining experiments we used a true random sample of games rather than a filtered random sample (including only games with inconsistent predictors) and the instructions accurately described the sampling procedure.

For each of the 29 games, participants were asked to predict which team won the game and to indicate their confidence in the prediction on a scale from 0 (not at all confident) to 10 (extremely confident). All participants were provided with information about the number of games won in the season by both teams and the point difference at halftime. The critical manipulation was that participants were either provided with the team names (e.g., New York Knicks vs. New Jersey Nets) or with labels for the team names (Team A vs. Team B). At the end of the questionnaire, all participants were asked whether they considered themselves basketball fans, whether they watched basketball on television, and if so, how often they watched per week. For the purposes of all analyses in this paper, a fan is defined as someone who considers him or herself a basketball fan and also watches basketball. The overall design was a 2 (Information: names vs. no-names) X 2 (Fan: fan vs. non-fan) between-subjects design.

Results and discussion

Neither the experimental condition nor the fan status reliably affected accuracy (Table 1). Participants in all conditions predicted at chance levels. However, it should be noted that it was possible to predict these games. The logistic regression model estimated from the data for the 1996/97 NBA season correctly predicted 62% of the games. Although fans were not more accurate than non-fans, fans were more confident (M = 5.72, SD = 1.08 vs. M = 4.99, SD = 1.36, respectively), $F(1, 76) = 8.42$, $p < .005$, $\eta^2 = .10$. Participants in the names condition (M = 5.61, SD = 1.34) were more confident than participants in the no-names condition (M = 5.12, SD = 1.16), $F(1, 76) = 4.04$, $p < .048$, $\eta^2 = .05$. These effects were qualified by an interaction, $F(1, 76) = 4.00$, $p < .049$, indicating that adding names to the game information did not increase the confidence of non-fans, $t < 1$, but reliably increased the confidence of fans, $t(76) = 2.87$, $p < .005$, $\eta^2 = .10$ (Table 1).

Experiment 2

Although we randomly sampled games in Experiment 1, the stimuli used were a filtered random sample because we selected only games with inconsistent predictors. As shown in the section on validity of statistical cues, these games are less frequent than games with consistent predictors, because teams with better records are likely to lead in the half. This is important because detrimental effects of knowledge of team names on accuracy are most likely in predictions of games with consistent predictors. In such situations, when names are not provided, the most reasonable strategy is to predict in line with the statistical odds. However, when team names are provided, participants might predict against the statistical odds if they have a particular theory about the teams. Such a strategy would lead to inconsistent predictions that, in the long run, will be less accurate. To test whether the knowledge of team names would decrease the accuracy of predictions, we used an unfiltered random sample of games in this experiment.

We also asked participants about their beliefs regarding how knowing the team names affected their predictions. We expected that although team names would reduce the accuracy of prediction, participants, in particular basketball fans, would believe that their knowledge of the teams improved their predictions.

Method

Participants. One hundred and twenty undergraduate students from Princeton University were recruited for a mass questionnaire session and were paid for their completion of the surveys. Participants were randomly assigned to one of the same two experimental conditions as in Experiment 1.

Game selection. For each of the 29 NBA teams, we randomly selected 1 game from the season previous to the study. If a game was repeatedly selected, it was replaced by a new game. For 20 of the 29 games, the team records and the score at halftime were consistent (e.g., Team A has a better record and is leading at the half), and the remaining games had inconsistent predictors.

Procedures. The procedures were the same as in Exp. 1. However, participants were asked additional questions at the end of the study. In the no-names condition, participants were asked whether they would have preferred to know the team names and whether knowing the team names would have improved their predictions. In the names condition, participants were asked whether they would have preferred not to know the team names and whether knowing the team names improved their predictions. Participants who responded in the negative to the latter question were asked whether not knowing the team names would have improved their predictions. The overall design was a 2 (Information) X 2 (Fan) between-subjects design.

Results and discussion

As predicted, participants in the names condition ($M = .63$, $SD = .10$) were less accurate than participants in the no-names condition ($M = .66$, $SD = .07$), $F(1, 116) = 4.38$, $p < .04$, $\eta^2 = .04$ (Table 1). Fans ($M = 5.54$, $SD = 1.33$) were more confident than non-fans ($M = 5.08$, $SD = 1.67$), but the effect did not reach significance, $F(1, 116) = 2.24$, $p = .14$, $\eta^2 = .02$.

Fans' intuitions about the role of the team names were inconsistent with the finding that providing the team names actually reduced the accuracy of predictions. In the no-names condition, 68.4% of the fans wanted to know the team names versus 19.0% of non-fans, $\chi^2(1) = 14.13$, $p < .001$. Moreover, 63.2% of the fans believed that knowing the team names would have improved their predictions versus 26.2% of non-fans, $\chi^2(1) = 7.61$, $p < .006$.

In the names condition, only 27.3% of the fans and 37.1% of the non-fans reported that they would have preferred to be ignorant of the names, $\chi^2 < 1$. When asked whether knowing the team names improved their predictions, 45.5% of the fans reported that it did so versus 11.4% of the non-fans, $\chi^2(1) = 8.44$, $p < .004$. It is worth pointing out that fans who believed that the names improved their predictions were both less accurate and more confident than fans who did not believe that the names helped them (.62 vs. .66 and 5.84 vs. 5.65 respectively). Although the majority of participants in the names condition did not think that the names improved their predictions, when probed further only 8.5% (5 participants: 1 fan and 4 non-fans) believed that the names might have decreased the accuracy of their predictions.

Experiment 3

The objective of Experiment 3 was to study one possible mechanism underlying the detrimental effect of knowledge of team names on predictions. We expected that participants would rely on the familiarity of the teams in their predictions despite the fact that the teams' records for the season are provided. To the extent that the familiarity of the teams is correlated with their perceived strength, participants should be more likely to choose familiar teams than unfamiliar teams even when the statistical odds favor the latter.

To manipulate familiarity experimentally, we created a familiarity index for each NBA team, using an independent sample of participants. The familiarity index correlated with the team record, but the correlation was modest, $r(28) = .35$, $p = .07$, excluding the record of the Chicago Bulls[3]. Based on the familiarity index, we selected games in which a familiar team was playing against less familiar teams with better season records, and games in which an unfamiliar team was playing against more familiar teams with worse records. We expected that in both cases participants in the names condition would be more likely to choose the familiar team than participants in the no-names condition. Finally, we randomly sampled games from the records of the remaining teams.

Method

Participants. One hundred and sixteen undergraduate students from Princeton University were recruited for a mass questionnaire session and paid for their completion of the surveys. Participants were randomly assigned to one of two experimental conditions. Another 30 participants were recruited for a study on familiarity of NBA teams.

Familiarity index. Participants were told they were taking part in a memory task and were asked to list as many NBA teams as they could recall in order that they came to mind. If participants could not remember the full names of the team (e.g., New York Knicks), they were allowed to provide either the team's city or the team name. The proportion of participants who listed a team name served as a familiarity index for this team. For example, 29 out of 30 participants listed Los Angeles Lakers – the most familiar team, with index of .97 – and only 13 participants listed Minnesota Timberwolves - familiarity index of .43.

Game selection. Although the Lakers were the most familiar team, several teams had a better record for the season: the Sacramento Kings, Dallas Mavericks, and San Antonio Spurs. For each of the three teams, we randomly selected 2 out of the 4 games played against the Lakers. The Detroit Pistons had as good a record as the Lakers and we added the 2 games played by both teams to the 6 selected games, giving us a total of 8 games.

The Timberwolves had a very good record for the season, but the team was not familiar to participants. We selected 4 teams that were more familiar but had worse records: the New Jersey Nets, Philadelphia 76ers, Boston Celtics, and Houston Rockets. We randomly selected 2 of the 4 games the Houston Rockets played against the Timberwolves and added the games played by the other teams against the Timberwolves (2 games for each team), giving us a total of 8 games.

The 16 selected games involved 10 teams. From the record of each of the remaining 19 NBA teams, we randomly sampled one game following the selection procedures of Exp. 2. The final sample of games consisted of 8 games of a team playing less familiar teams with better season records, 8 games of a team playing more familiar teams with worse records, and 19 randomly sampled games.

Procedures. The procedures were the same as in Exp. 1. The overall design was a 2 (Information) X 2 (Fan) between-subjects design.

Results and discussion

Although the Lakers were competing against teams with better records for the season, participants in the names condition ($M = .39$, $SD = .27$) were more likely to predict that the Lakers would win than participants in the no-names condition ($M = .21$, $SD = .12$), $F(1, 112) = 30.59$, $p < .001$, $\eta^2 = .21$. Fans ($M = .38$, $SD = .26$) were also more likely to choose the Lakers than non-fans ($M = .24$, $SD = .19$), $F(1, 112) = 14.19$, $p < .001$, $\eta^2 = .11$. These effects were qualified by an interaction, $F(1, 112) = 4.97$, $p < .028$, indicating that the tendency to select the Lakers was especially pronounced for fans in the names condition (Table 2). The analysis of confidence and accuracy showed that fans ($M = 5.82$, $SD = 1.64$) were more confident than non-fans ($M = 4.91$, $SD = 1.66$), $F(1, 112) = 8.56$, $p < .004$, $\eta^2 = .07$ and less accurate ($M = .60$, $SD = .12$ vs. $M = .56$, $SD = .16$, respectively), although the latter effect did not reach significance, $F(1, 112) = 2.73$, $p = .102$.

The findings for the Timberwolves mirrored the findings for the Lakers. Although the Timberwolves were competing against teams with worse records, participants in the names condition ($M = .43$, $SD = .14$) were less likely to predict that the Timberwolves would win than participants in the no-names condition ($M = .50$, $SD = .14$), $F(1, 112) = 6.98$, $p < .009$, $\eta^2 = .06$. This effect seemed more pronounced for fans (Table 2), but the interaction was not significant, $F < 1$. The analysis of confidence and accuracy showed that fans ($M = 5.91$, $SD = 1.33$) were more confident than non-fans ($M = 4.99$, $SD = 1.72$), $F(1, 112) = 9.50$, $p < .003$, $\eta^2 = .08$, although they were not more accurate, $F < 1$.

As shown in Table 1, the analysis of the predictions of the 19 randomly sampled games replicated Experiment 2. Participants in the names condition ($M = .78$, $SD = .13$) were less accurate than participants in the no-names condition ($M = .83$, $SD = .08$), $F(1, 112) = 3.60$, $p <$

.060, $\eta^2 = .03$. Although fans were not more accurate than non-fans, $F < 1$, they were more confident ($M = 6.21$, $SD = 1.42$ vs. $M = 5.58$, $SD = 1.63$, respectively), $F(1, 112) = 4.56$, $p < .035$, $\eta^2 = .04$.

Experiment 4

Experiments 2 and 3 showed that knowledge of the team names reduced the accuracy of predictions. In this experiment, we tested whether this knowledge can also affect the betting behavior of participants and, ultimately, their earnings. In this experiment, participants made predictions in a context where real money was at stake. Basketball fans predicted the outcomes of a set of 58 randomly sampled games, and wagered money on each prediction. They could wager 10, 20, or 30 cents on each of the games.

If the reduced accuracy of predictions in the name condition, coupled with the overconfidence of these predictions, translates into actual financial losses, fans provided with the team names should earn significantly less than their counterparts who are not privy to the extra information. In this experiment, participants were also provided with feedback about the accuracy of their predictions after each game. We provided feedback because this closely approximates realistic conditions of betting. Moreover, to the extent that participants learn from this feedback, they can adjust their predictions and, presumably, reduce the influence of the name information on their predictions. However, to the extent that they believe that the names are helping their predictions, the feedback information may not be sufficient to overcome the detrimental effect of names on the accuracy of prediction.

Method

Participants. Twenty participants from Princeton University and the University community volunteered for the study. Participants were informed that they would earn money (with a possible range from \$4 to \$21.40) based upon their knowledge of the NBA. There was a base payment of \$4.00, with money added depending on accuracy of prediction and amount wagered. All participants reported being basketball fans, who followed the NBA and watched basketball on television. They were randomly assigned to one of two experimental conditions.

Game selection. For each of the 29 NBA teams, 2 games were randomly sampled for each team from the previous season. If a duplicate game was sampled, it was randomly replaced by a new game from that team. Of the 58 games sampled, 37 of them had consistent predictors, 18 had inconsistent predictors and 3 were unclassifiable (halftime score tied).

Procedure. Participants were informed that each NBA team plays 82 games in a season, and that all of the games to be predicted were randomly sampled from the previous season. Using a computer program, games were presented in a random order, and for each game, the participants predicted the winner and bet 10, 20 or 30 cents on the prediction. After each trial, they were immediately given feedback as to whether or not their prediction was accurate. Based upon their overall performance, participants earned or lost the amount they wagered. The same information manipulation was used as in the previous studies: all participants received both the score at halftime and winning record of each team. Half of the participants were randomly assigned to the name condition, in which they also received the names of the teams. Finally, after making all of the predictions, participants completed a questionnaire in which they assessed the usefulness of the information in making their judgments.

Results and discussion

As expected, participants predicting with only the statistical information earned more (M = \$2.78, SD = .91) than those with the team name information (M = \$1.64, SD = 1.17), $t(19) = 2.35$, $p < .03$, $\eta^2 = .22$. This difference was driven by the fact that the participants in the names condition were less accurate (*M* = .56, SD = .09) than those without the names (*M* = .61, SD = .11), $t(19) = 3.97$, $p < .02$, $\eta^2 = .45$. The Sobel test for mediation of accuracy on money earned was significant, $t(19) = 2.22$, $p < .03$. There was no difference in the bet amount across the two conditions (*M* = .23 vs. *M* = .24, $t < 1$).

As in Experiment 2, fans' beliefs about the usefulness of the team information contrasted with the experimental findings. Ninety percent of those with the names (that is all but 1 participant) believed that the knowledge of the names improved their predictions. Similarly, 70% of participants without the team names said they would have preferred to know the names. This belief that the team names improve predictions can cause the bias to persist, even when participants are provided with continuous feedback about their performance. In fact, when comparing the accuracy between the predictions for the first 29 games and the predictions for the second 29 games in the names condition, there was no difference (*M* = .57 vs. *M* = .56, respectively, $t < 1$, $p = .99$).

These data provide converging evidence for the phenomenon demonstrated in the previous studies, and extend them into a more realistic situation. With real money at stake, basketball fans still displayed decreased accuracy with the additional information. The participants who were privy to the name information earned roughly 60% less than those predicting without that information, and the majority of participants reported that this information improved, or *would have* improved, their performance at the task. Although the name effect seems small in terms of accuracy alone, this study demonstrates how the consequences can be magnified in a real life situation. Furthermore, the bias persisted even in the presence of feedback information that allowed for potential learning and adjustment of cue use.

Analyses across experiments

Because all participants were drawn from the same population, we report additional analyses across experiments to address possible mechanisms leading to reduced accuracy and increased confidence. Participants' accuracy and confidence were uncorrelated (r = .03 across the first 3 experiments), suggesting distinct determinants of confidence and accuracy. In fact, providing the team names increased the confidence of fans only, but reduced the accuracy of both fans and non-fans.

Determinants of confidence. In the first 3 experiments, providing the team names increased fans' confidence but either had no effect or decreased non-fans' confidence (Table 1). This pattern is consistent with the hypothesis that team names cue retrieval of specific knowledge for fans and that this knowledge increases their confidence. However, the interaction of information and fan status was not significant in Experiments 2 and 3, most likely because of lack of statistical power. The confidence judgments were submitted to a 2 (Information) X 2 (Fan status) X 3 (Experiment) ANOVA. Fans were more confident than non-fans, $F(1, 304) = 14.87$, $p < .001$, $\eta^2 = .05$. More important, the only other significant effect was the interaction of fan status and information, $F(1, 304) = 4.72$, $p < .031$, $\eta^2 = .02$. As shown in Fig. 1, whereas the confidence of fans increased when they were provided with the team names, $t(312) = 2.10$, $p < .036$, the confidence of non-fans was not affected, $t < 1$. Fans were also more confident than non-fans in the names condition, $t(312) = 4.33$, $p < .001$, but not in the no-names condition, $t(312) = 1.26$, $p = .21$.

Correlational analyses are also consistent with the hypothesis that the team names cue specific knowledge and the retrieval of this knowledge increases confidence. In all experiments, we asked participants to report how frequently they watch basketball on television. To the extent that higher frequency reports reflect better knowledge of the teams, or even incorrect beliefs of better knowledge, the frequency reports should correlate with confidence when the names are provided. In fact, as shown in Table 3, in the first 3 experiments, the correlation between confidence and frequency of watching basketball games was higher when team names were provided than when they were not provided. To test whether the difference between these correlations was significant, we regressed the participants' confidence on the information condition, the frequency of watching basketball on television, their interaction, and two dummy variables controlling for the experiments. As expected, the interaction of information condition and frequency of watching basketball was significant, $t(310) = 2.84$, $p < .005$, indicating that the difference in correlations was statistically significant.

Determinants of accuracy: The reduced salience of statistical information. If knowledge of the team names reduces the salience of the statistical information, participants should be less likely to rely on this information when making their predictions. In turn, this should reduce the accuracy of predictions. To test this prediction, for each participant in Experiments 2 and 3, we computed the biserial correlation between the participant's prediction and each statistical cue across all games: the difference in the score at halftime and the difference in the team records for each game. The average correlation between the score at halftime and the participants' predictions was .53 ($SD = .22$). The average correlation between the difference in team records and the participants' predictions was .52 ($SD = .23$). More important, we submitted the Fisher z-transformation of these correlations[4] to a 2 (Statistical cue: score vs. record) X 2 (Information) X 2 (Fan status) X 2 (Experiment) mixed subjects ANOVA with the first factor as a repeated measure. As shown in Fig. 2, participants in the names condition were less likely to rely on the statistical cues than participants in the no-names condition. This was the only significant effect, $F(1, 228) = 14.92$, $p < .001$, $\eta^2 = .06$.

The accuracy of participants' predictions was highly correlated with the correlation between their predictions and the difference in the score at halftime, $r(236) = .70$, $p < .001$, and modestly correlated with the correlation between their predictions and the difference in team records, $r(236) = .30$, $p < .001$. In other words, the more participants relied on the statistical cues the more accurate they were. It is important to show that the decreased reliance on statistical information in the names condition (Fig. 2) reduced the accuracy of predictions. To demonstrate that the reliance on statistical information mediated the effect of information on the participants' accuracy, we regressed accuracy on the information condition, and on the biserial correlations of the participants' predictions with the statistical cues. As shown in Table 4, knowledge of team names was a significant predictor of accuracy only when it was the only predictor (Model 1). Entering the biserial correlations between the participants' predictions and the statistical cues completely removed the effect of the team names (Model 4). In fact, entering the biserial correlation between predictions and score at halftime was sufficient to eliminate the knowledge effect on accuracy (Model 2). The Sobel test for mediation was significant, $t(236) = 3.45$, $p < .001$. Entering the biserial correlations of predictions with difference in team records reduced the effect of information condition, although it did not completely eliminate it (Model 3). However, even in this model, the Sobel test for mediation was significant, $t(236) = 2.31$, $p < .022$.

Determinants of accuracy: The role of familiarity. Reducing the salience of the statistical information reduced the accuracy of predictions. It is possible that this reduction in accuracy can be entirely attributed to increased noise or random error in the predictions. The alternative is that knowledge of the team names introduces systematic biases in predictions. Experiment 3 identified one such bias. Participants used the familiarity of the teams to make predictions even when this familiarity was inconsistent with the team records. This finding was obtained on a selected sample of games. To test whether the finding holds across all games, we computed the proportion of times each participant bet on the more familiar team and submitted these proportions to a 2 (Information) X 3 (Experiment) ANOVA (see footnote 3 for the Chicago games). As expected, when the team names were provided, participants were more likely to bet on the more familiar team ($M = .53$, $SD = .09$) than when the names were not provided ($M = .48$, $SD = .06$), $F(1, 310) = 29.45$, $p < .001$, other Fs < 1.

In the case of games with consistent statistics (team record and halftime score), this reliance on familiarity can be particularly detrimental because participants can predict against the statistical odds. As shown in the section on validity of statistical cues, when the statistics were consistent, a judgment of the leading team winning the game was correct 86.5% of the time. Thus, any time participants are choosing against the odds, they are dramatically reducing their chances of correct prediction. There were 23 games (11 in Experiment 2 and 12 in Experiment 3) in which the statistical odds and the familiarity of the team names were inconsistent. A 2 (Information) X 2 (Experiment) ANOVA on the proportion of times each participant choose the more familiar team showed that participants in the names condition ($M = .19$, $SD = .22$) were more likely to choose the more familiar team against the statistical odds than participants in the no-names condition ($M = .07$, $SD = .11$), $F(1, 232) = 29.11$, $p < .001$, other Fs < 1. Thus, when provided with the team names, participants were more than twice more likely to predict against the statistical odds. Assuming a ceiling of accuracy of 85% when the statistical cues are consistent and 20% rate of predictions against the statistical odds, the expected accuracy rate is 71%. In fact, the overall accuracy rate for consistent games across Experiments 2 and 3 was 70% in the names condition.

The systematic bias in predictions introduced by knowledge of the team names is also confirmed by analyses at the level of the games (where the unit of analysis is the game rather than the participant). In the clear majority of games, participants were more likely to bet on the statistical odds, indicating that they did not ignore the statistical information. In fact, the predictions, aggregated across participants, in the names condition were highly correlated with the predictions in the no-names condition. For all experiments, the correlation was above .92. However, the question is whether the variance remaining after removing the variance due to the statistical information is systematically related to team familiarity. To test this hypothesis, we regressed the predictions in the names condition on the predictions in the no-names condition and correlated the regression residuals with the difference in team familiarity. As shown in Fig. 3 and Table 5, these residuals were systematically related to the difference in team familiarity. That is, after removing the influence of the statistical information, the predictions were systematically biased by differences in the familiarity of the teams.

General Discussion

In a series of studies, we explored how people predict uncertain outcomes in a well-defined domain. Participants used statistical information appropriately and, in fact, the weight of this information in their predictions was similar to the predictive validities of the statistical cues. However, when provided with a superfluous knowledge cue in addition to the statistical information, participants' judgments deteriorated. In four experiments, we showed that this additional knowledge can increase confidence in predictions and at the same time decrease their accuracy.

In Experiment 1, we selected basketball games in which knowledge of the team names could have improved accuracy. In all games, the statistical cues were inconsistent with one another and the additional information provided by the team names could have reduced subjective uncertainty. However, the only effect of the additional information was to increase the confidence of basketball fans.

In Experiments 2 and 4, participants who were provided with team names were less accurate in predicting the outcomes of a random sample of NBA games than participants who were not provided with this information. At the same time, participants' theories about the value of the knowledge of team names were at odds with this finding. When not provided with the team names, the majority of basketball fans wanted to know the names and believed that this knowledge would have improved their predictions. When provided with the names, in Experiment 2 close to 50% of fans believed that the names improved their predictions, and all but one of the remaining participants believed that the names did not make a difference. In Experiment 4, in which fans were financially motivated to perform well, 90% believed that the names improved their predictions. Although fewer non-fans (in Experiment 2) wanted to know the team names and thought that this knowledge would have improved their predictions, once provided with the names only about one-third preferred not to know the names and only about 10% thought that the names could have worsened their predictions.

Experiment 4 also showed that the decreased accuracy in prediction translates into financial losses in a real betting situation. When wagering money on their predictions, basketball fans earned significantly less money when making forecasts with the team name information. In addition, although fans were provided with feedback after each of their predictions, they did not seem to adjust their use of the available cues to avoid the bias. This finding seems less surprising in the context of their beliefs that the knowledge of team names improved their predictions.

In Experiment 3, we showed that the additional knowledge did not simply introduce random noise in the judgments making the statistical information less salient, but instead introduced systematic biases in predictions. We identified one possible mechanism contributing to the detrimental effects of additional knowledge on predictions, namely participants' reliance on the familiarity of the teams. Although familiarity is modestly correlated with the success of the teams, it is an imperfect predictor and often the team record and the familiarity of the teams can be inconsistent. Moreover, in the current studies, participants were provided with the team records for the season. Despite the availability of the statistical cues, participants were more likely to predict that the more familiar team would win when provided with the names than when not provided with the names. This was the case even when both statistics clearly favored the less familiar team. Presumably, participants overestimated the chances of winning of recognized or more familiar teams. This would be consistent with the recognition heuristic (Goldstein & Gigerenzer, 2002; Schooler & Hertwig, 2005).

It should be noted that the measure of familiarity of the teams was far from perfect. First, it was not adjusted to the specific knowledge of individual participants. For example, the familiarity of the teams will most likely vary for participants from the East and the West coasts. Second, even if the familiarity of the teams is correlated with the perceived strength of the teams, it is also affected by a variety of other factors. However, despite all these problems, the familiarity of the teams, obtained from an independent sample of participants, significantly predicted the performance of the participants. Future studies should explore the exact substantive biases introduced by additional 'non-statistical' knowledge in predictions of uncertain outcomes. The goal of Experiment 3 was to show that team names introduce systematic biases in predictions.

Less accurate but more overconfident predictions

Ironically, although knowledge of team names reduced the accuracy of basketball fans, it also increased their confidence. We suggest that this increase in confidence results from the retrieval of specific knowledge about the teams. Two types of evidence are consistent with this hypothesis. First, the confidence of non-fans who presumably do not have as rich knowledge of the teams as fans was not affected by the provision of the names. In contrast, the confidence of fans increased when they were provided with the names. For fans, the names could have activated specific knowledge about the teams, knowledge that they believe is improving their predictions. Second, the frequency of watching basketball on television predicted confidence in predictions only when the team names were provided. When the names are not provided, participants cannot retrieve specific knowledge about the teams or the games (an NBA season has more than 1000 games). At best, participants who watch basketball frequently can activate knowledge relevant to statistical prediction. In contrast, when the names are provided, participants can retrieve specific knowledge about the teams and to the extent that participants who watch basketball more frequently know more about the teams they should be able to retrieve more knowledge. Our findings are consistent with prior studies showing that additional information often results in increased confidence without corresponding effects on accuracy (Arkes et al., 1986; Gill et al., 1998; Oskamp, 1965; Stewart et al., 1992).

In fact, knowledge of the teams reduced the accuracy of prediction. Our findings suggest that, when provided with the names, participants treated the games as ‘unique’ events relying on idiosyncratic knowledge of the teams in their predictions. In general, prediction strategies that focus on the teams or knowledge outside the relevant statistics should lead to less accurate predictions. Halberstadt and Levine (1999) found that predictions of the outcomes of college basketball games were less accurate when basketball fans analyzed the reasons why one of the teams could win (e.g., “They have a huge center.”). This task involves retrieval of specific knowledge about the teams and, most likely, leads to ignoring relevant statistical information (e.g., ranking of the teams). In our studies, detailed analyses of the participants’ predictions showed that the additional information reduced the weight of the statistical cues in predictions and, ultimately, reduced their accuracy. Moreover, additional analyses showed that this information did not simply introduce random error in the predictions but systematically biased them.

The illusion of knowledge

In certain situations, people may be better off consulting fewer pieces of information when making decisions. Specifically, when an uncertain event can be statistically characterized, it would be best not to consult any “event-specific” information. However, this runs against the intuitions of people knowledgeable about the particular domain under consideration. In our study, the majority of basketball fans believed that knowledge of the names would improve their predictions. Not surprisingly, when given these names, fans became more confident. This constellation of factors – additional information is believed to improve the decision but its effect is to reduce accuracy – produces “the illusion of knowledge” effect. Because this effect is based on the powerful intuition that more knowledge is beneficial for decisions, it may be very difficult to dissuade people from using “event-specific” information. This is one of the reasons why after decades of repeated demonstrations that actuarial predictions are better than clinical predictions, people still rely on the latter (Dawes et al., 1989). Moreover, even when presented with the statistical evidence, people persist in their beliefs. The belief in the “hot hand” (making a successful basketball shot makes the success of subsequent shots more likely) is one such example (Gilovich, Vallone, & Tversky, 1985). Despite the lack of evidence for this phenomenon, it is nearly impossible to convince basketball fans (unless they are dedicated students of judgment and decision making) of its unreality.

The illusion of knowledge is related to at least three other phenomena: the hindsight bias (Fischhoff, 1975), the curse of knowledge (Camerer, Loewenstein, & Weber, 1989), and the illusion of transparency (Gilovich, Savitsky, & Medvec, 1998). In the hindsight bias, knowledge of outcomes biases retrospective predictions by increasing their congruence with the outcome. In the curse of knowledge, people with more knowledge are unable to predict the knowledge of people with less knowledge, generally overestimating the knowledge of the latter. In the illusion of transparency, people overestimate the ability of others to read their internal states. In all four phenomena, more knowledge systematically biases judgments. However, there are also important differences between these phenomena. For example, in the hindsight bias, the curse of knowledge, and the illusion of transparency, people insufficiently discount the impact of their own knowledge on judgments. In the illusion of knowledge, insufficient discounting does not seem to be involved. Future studies should identify whether these phenomena share common mechanisms through which knowledge reduces the accuracy of judgments.

Illusions of knowledge can be pervasive. Heath and Tversky (1991) showed that people were paying a 20% premium to bet on areas familiar to them. Ultimately, because of the overconfidence in these areas, they were most likely to lose money. Similarly, Weber, Siebenmorgen, & Weber (2005) have shown that when participants were given the names of financial assets they perceived these assets as less risky and expected greater return. In fact, participants felt more competent when evaluating stocks with more familiar names. People also exhibit a “home bias” (allocating more funds to domestic vs. foreign assets) and asset name familiarity bias (allocating more funds to stocks with greater name recognition) (Boyd, 2001; Lewis, 1999; Huberman, 2001). The biases in all of these cases, from the prediction of basketball games to the allocation of assets, can be traced to illusions of knowledge.

References

Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes, 37*, 93-110.

Bastardi, A., & Shafir, E. (1998). On the pursuit of useless information. *Journal of Personality and Social Psychology, 75(1)*, 19-32.

Boyd, M. (2001). On ignorance, intuition, and investing: A bear market test of the recognition heuristic. *The Journal of Psychology and Financial Markets, 2(3)*, 150-156.

Camerer, C., Loewenstein, G., & Weber, M. (1989). The curse of knowledge in economic settings: An experimental analysis. *Journal of Political Economy, 97*, 1232-1254.

Cooper, H., DeNeve, K. M., & Mosteller, F. (1992). Predicting professional sports game outcomes from intermediate game scores. *Chance: New Directions for Statistics and Computing, 5(3-4)*, 18-22.

Dawes, R.M. (1971). A case study for graduate admissions: Application of three principles of human decision making. *American Psychologist*, *26*, 180-188.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668-1674.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment, 50(3)*, 387-395.

Fischhoff, B. (1975). Hindsight ≠ Foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288-299.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98(4)*, 506-528.

Gill, M. J., Swann, W. B. Jr., & Silvera, D. H. (1998). On the genesis of confidence. *Journal of Personality and Social Psychology, 75*, 1101-1114.

Gilovich, T., Savitsky, K., & Medvec, V. H. (1998). The illusion of transparency: biased assessments of others' ability to read one's emotional states. *Journal of Personality and Social Psychology, 75(2)*, 332-346.

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

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

Gigerenzer, G., & Goldstein, D. G. (1999). Betting on one good reason: The take the best heuristic. In G. Gigerenzer & P. Todd (Eds.), *Simple heuristics that make us smart.* (pp. 75-95). New York: Oxford University Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. New York: Oxford University Press.

Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin, 73(6),* 422-432.

Goldstein, D. G. & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer & P. Todd (Eds.), *Simple heuristics that make us smart.* (pp. 37-58). New York: Oxford University Press.

Goldstein, D. G. & Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review, 109(1),* 75-90.

Halberstadt, J. B., & Levine, G. M. (1999). Effects of reasons analysis on the accuracy of predicting basketball games. *Journal of Applied Social Psychology, 29(3)*, 517-530.

Hammond, K. R. (1996). *Human Judgment and Social Policy*. New York: Oxford University Press.

Heath, C., & Tversky, A. (1991). Preference and belief: Ambiguity and competence in choice under uncertainty. *Journal of Risk and Uncertainty*, *4(1),* 5-28.

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

Hilton, J. L., & Fein, S. (1989). The role of typical diagnosticity in stereotype-based judgments. *Journal of Personality and Social Psychology, 57*, 201-211.

Huberman, G. (2001). Familiarity breeds investment. *The Review of Financial Studies, 14(3)*, 659-680.

Iyengar, S. S. & Lepper, M. R. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology. 79*, 995-1006.

Kahneman, D. & Lovallo, D. (1993). Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking. *Management Science, 39(1),* 17-31.

Lewis, K. K. (1999). Trying to explain home bias in equities and consumption. *Journal of Economic Literature, 37*, 571-608.

Nisbett, R., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology, 13*, 248-277.

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

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29(3)*, 261-265.

Schooler, L. J. & Hertwig, R. (2005). How forgetting aids heuristic influence. *Psychological Review, 112(3), 610-628.*

Schwartz, B. (2004). *The paradox of choice: Why more is less*. New York: HarperCollins.

Stewart, T. R., Heideman, K. F., Moninger, W. R., & Reagan-Cirincione, P. (1992). Effects of improved information on the components of skill in weather forecasting. *Organizational Behavior and Human Decision Processes, 53*, 107-134.

Todorov, A. (1997). Another look at reasoning experiments: Rationality, normative models, and conversational factors. *Journal for the Theory of Social Behaviour, 27*, 387-417.

Todorov, A. (2002). *Predicting real outcomes: When heuristics are as smart as statistical models*. New York University. Unpublished manuscript.

Weber, E. U., Siebenmorgen, N., & Weber, M. (2005). Communicating asset risk: How name recognition and the format of historic volatility information affect risk perception and investment decisions. *Risk Analysis*, *25*, 597-609.

Zukier, H. (1982). The dilution effect: The role of the correlation and the dispersion of predictor variables in the use of nondiagnostic information. *Journal of Personality and Social Psychology, 43*, 1163-1174.

Footnotes

1. This effect is similar to the dilution effect (Nisbett et al., 1981; Zukier, 1982), where non-diagnostic or pseudo-diagnostic information (e.g., Hilton & Fein, 1989) reduces the influence of actuarial information. However, this effect has been interpreted to have different implications for the optimality of judgments. Because people make insufficiently regressive predictions when provided with diagnostic information, adding pseudo-diagnostic information improves these predictions by making them more regressive, although people do not truly understand the regression principle.

2. The logistic regression model was estimated from an analysis of all games, not only the inconsistent games. The accuracy rate for this model was practically identical to the rate obtained from a discrimination analysis model (65.6%) and slightly better than the rate of a Bayesian model (60.9%). Todorov (2002) also showed that a version of the Take-the-Best algorithm (Gigerenzer & Goldstein, 1996, 1999) predicts 60.4% of the inconsistent games.

3. Including this record, the correlation was .25. The Chicago Bulls were the second most familiar team after Los Angeles Lakers. However, we believe this familiarity was due to the recent history of the team with Michael Jordan. At the time of the experiments, the Chicago Bulls were not very successful and this fact was known to most participants. For these reasons, the analysis of familiarity reported below also excluded all Chicago games (3 in Experiment 2, and 2 in Experiment 3), although including these games did not change the results.

4. This is a standard transformation used to avoid skewed correlation distributions. We conducted the same analyses on the raw correlations and the results were identical.

Table 1. Proportion (Standard Deviations) of Accurate Predictions and Mean Confidence (Standard Deviations) in Predictions of Randomly Sampled Basketball Games as a Function of Information and Fan Status.

	Information			
	Statistical		Statistical + team names	
Experiment 1	Non-fans	Fans	Non-fans	Fans
Accuracy	.49 (.08)	.53 (.07)	.52 (.08)	.51 (.09)
Confidence	4.99 (1.34)	5.23 (1.02)	5.00 (1.41)	6.29 (0.86)
Experiment 2				
Accuracy	.66 (.06)	.67 (.07)	.62 (.11)	.64 (.08)
Confidence	5.16 (1.22)	5.32 (0.96)	4.99 (2.07)	5.74 (1.58)
Experiment 3				
Accuracy	.83 (.08)	.82 (.08)	.78 (.14)	.80 (.11)
Confidence	5.74 (1.64)	6.10 (1.48)	5.42 (1.63)	6.33 (1.37)
Experiment 4				
Accuracy		.61 (.11)		.56 (.09)

Note. Predictions were based on 29 games in Experiment 1, 29 games in Experiment 2, a subset of 19 games in Experiment 3, and 58 games in Experiment 4. All games in Experiment 1 were with inconsistent predictors (e.g., the team with better record is losing at the half).

Table 2. Proportion (Standard Deviations) of Bets, Accuracy (Standard Deviations) of Predictions, and Confidence (Standard Deviations) in Predictions of Los Angeles Lakers and Minnesota Timberwolves Games as a Function of Information and Fan Status (Experiment 3).

	Information			
	Statistical		Statistical + team names	
LA Lakers games	Non-fans	Fans	Non-fans	Fans
Betting on LA Lakers	.18 (.11)	.24 (.12)	.31 (.23)	.53 (.28)
Accuracy	.62 (.11)	.57 (.17)	.59 (.13)	.55 (.15)
Confidence	5.03 (1.78)	5.61 (1.63)	4.80 (1.55)	6.06 (1.65)
MT Timberwolves games				
Betting on MT Timberwolves	.51 (.15)	.49 (.14)	.45 (.12)	.40 (.18)
Accuracy	.73 (.17)	.69 (.16)	.72 (.13)	.71 (.11)
Confidence	5.14 (1.91)	5.76 (1.30)	4.84 (1.53)	6.08 (1.37)

Note. The LA Lakers were competing against less familiar teams but with better records for the season. The MT Timberwolves were competing against more familiar teams but with worse records for the season.

Table 3. Correlations between Self-reports of Frequency of Watching Basketball Games on Television and Confidence in Predictions of Basketball Games as a Function of Information.

	Information	
	Statistical	Statistical + team names
Experiment 1	-.04	.54***
Experiment 2	.04	.26*
Experiment 3	.23	.38**
Across all experiments	.11	.36***

* $p < .05$

** $p < .01$

*** $p < .001$

Table 4. Standardized Regression Coefficients of Information Condition, Correlations between Individual Predictions and Statistical Cues as Predictors of Accuracy (Experiments 2 and 3).

	Model 1	Model 2	Model 3	Model 4
Information condition	-.16, $p = .014$	-.01, $p = .91$	-.12, $p = .076$	.01, $p = .85$
Correlation with score at halftime		.68, $p < .001$		.66, $p < .001$
Correlation with team record			.22, $p = .001$	.09, $p = .058$

Note. The no-names condition was coded as 0 and the names condition was coded as 1. The biserial correlations between individual predictions and statistical cues were Fisher z-transformed for the regression analyses (see footnote 3).

Table 5. Correlations between Difference in Familiarity of Teams and Residuals of Predictions in the Names Condition after Controlling for Predictions in the No-names Condition.

	Correlation
Experiment 1	.57**
Experiment 2	.47*
Experiment 3	.66**
Across all experiments	.58***

* $p < .02$

** $p < .01$

*** $p < .001$

Figure legends

Figure 1. Confidence in predictions as a function of fan status and information condition. Confidence was measured on an 11-point scale ranging from 0 (not at all confident) to 10 (extremely confident). Data from the first three experiments. Error bars represent standard errors.

Figure 2. Biserial correlations between participants' predictions and statistical cues as a function of information condition. Data from Experiments 2 and 3. Error bars represent standard errors.

Figure 3. Scatter plot of residuals from regressing the aggregated predictions across participants in the names condition on the aggregated predictions across participants in the no-names condition and differences between the familiarities of the teams. Each point represents a game. Data from the first three experiments.

Figure 1

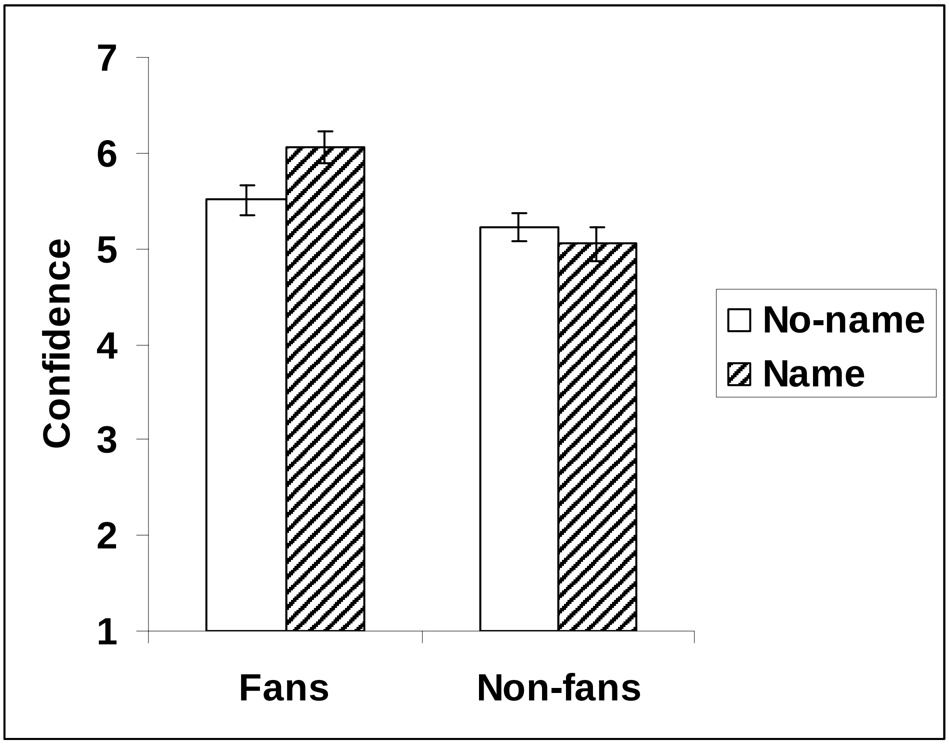


Figure 2

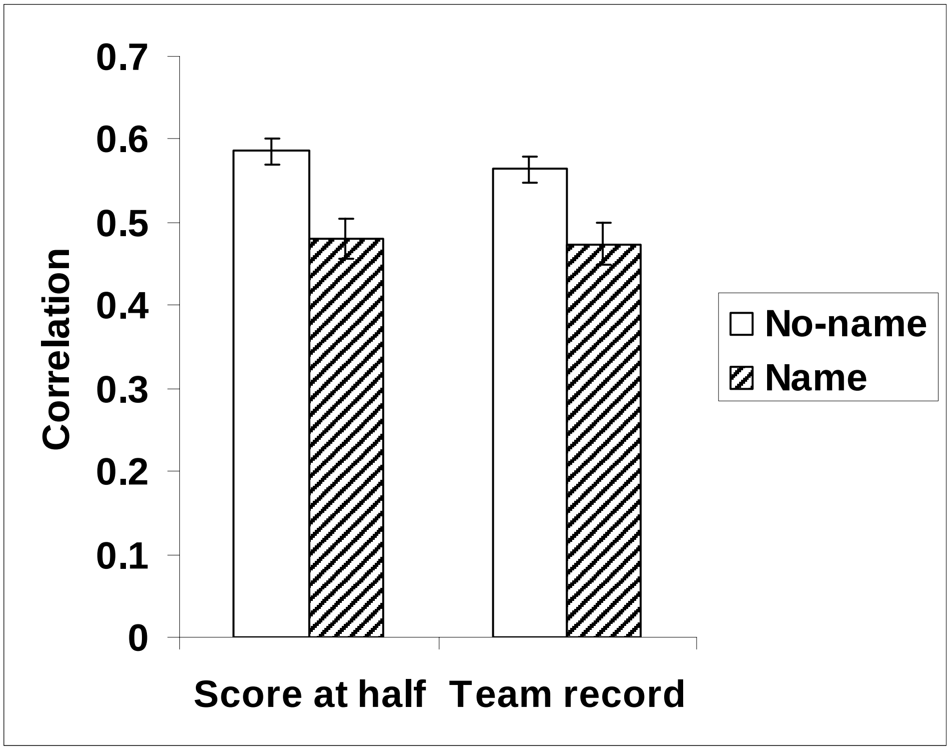
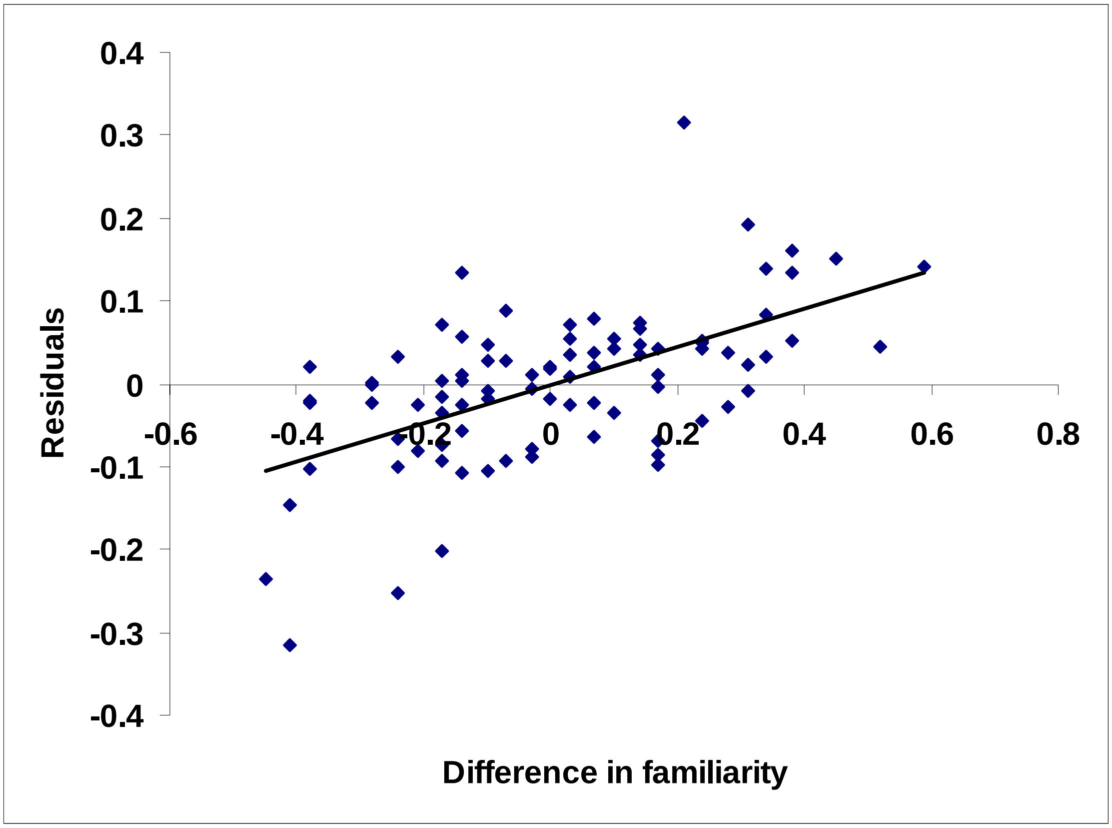

0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Correlation
Score at half
Team record
No-name
Name

Figure 3


0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
Residuals
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
Difference in familiarity