CORRECTED JANUARY 18, 2008; SEE LAST PAGE

Psychological Review
2004, Vol. 111, No. 3, 770–780

0033-295X/04/$12.00 DOI: 10.1037/0033-295X.111.3.770

The Hazards of Underspecified Models: The Case of Symmetry in Everyday Predictions

Peter Sedlmeier
Chemnitz University of Technology

Berna Kılınç
Boğaziçi University

Should one be more confident when predicting the whole (or an event based on a larger sample) from the part (or an event based on a smaller sample) than when predicting the reverse? The relevant literature on judgment under uncertainty argues that such predictions are symmetrical but that, as an empirical matter, people often fail to appreciate this symmetry. The authors show that symmetry in prediction does not necessarily hold. In addition to an empirical study involving predictions about soccer games, they develop a theoretical model showing that, at least for the ranges of numerical values usually found in everyday judgment problems, symmetry in predictions is uncommon when 2 different sample sizes are involved. The complexity of the theoretical model used in this analysis raises questions about model specification in judgmental research.

Which of the following assertions is more likely: that the candidate for the American presidency who wins the primary in New Hampshire will become the party nominee or that the party nominee has won the primary in New Hampshire? Recent research has argued that in situations such as this, one assertion is exactly as likely as the other. We argue instead that cases in which such symmetry holds, that is, cases in which predictions in both directions have equal probabilities, are rare. The prediction of an outcome of an extended process from a partial outcome does not necessarily amount to the prediction of a partial outcome from the outcome of the more extended process. One of the reasons why symmetry fails is the existence of ties, a possibility that cannot be easily ruled out in a game situation. But, symmetry may fail even when ties are ruled out, as we show below.

Although our example is a modest one, our argument highlights an issue that has been largely ignored by cognitive psychologists, namely, model specification. We contend that even when a theoretical framework—deriving from logic, probability theory, or statistics—is considered appropriate to analyze a real life situation, it might still be difficult to specify the normative model that applies to this situation. Our concern in this article differs from those expressed in the recent criticisms that question the existence of a unique theoretical framework or the extent to which the underlying assumptions about the problems given to participants are shared equally by experimenters and participants (Birnbaum, 1983; Cohen, 1981; Gigerenzer et al., 1989; MacDonald, 1986). That is, we assume in this article that there is no controversy about which theory of probability or statistics is adopted and that experimenters and participants have a common and clear understanding of the concept of probability and of the problem as stated. We argue that even under such conditions, the description of the problem may leave open alternative ways of modeling the problem situation, thus allowing for defensible alternative responses. It may then be very difficult or even practically impossible to obtain all the information necessary for a complete specification of a normative model. Our goal is to caution cognitive psychologists against the hazards of a too easy reliance on simple normative models as a standard for studying human rationality.

Previous Research

Whether prediction symmetry holds depends on the conditional probabilities used to quantify the strength of predictions that varies with sample size. A basic problem scheme that has been used by cognitive psychologists to test human cognitive adeptness in this regard is the following: Should one predict with more confidence from a single event to an aggregate of events or from an aggregate to a single event, or does each prediction carry the same weight? It has been repeatedly stated in the literature that the correct response to such problems is the third, symmetric alternative—that is, that sample size does not make a difference in the probabilities that should be assigned to predictions. We argue, instead, that *prediction symmetry* is seldom encountered in everyday problems involving unequal sample sizes.

Here are two examples gleaned from the literature to illustrate the question of prediction symmetry. The first example, what we call the *decathlon problem*, is drawn from Tversky and Kahneman (1982, Problem 5, p. 120):

> Which of the following events is more probable?
>
> (a) That an athlete won the decathlon, if he won the first event in the decathlon.

This research was supported by a Feodor Lynen Stipend of the Humboldt Foundation as well as a Habilitationsstipendium of the Deutsche Forschungsgemeinschaft to Peter Sedlmeier and by Boğaziçi Üniversitesi Research Funds (BAP 03HB202) to Berna Kılınç.

We are especially grateful to Duncan Luce for valuable advice. We also thank Valerie Chase, Gerd Gigerenzer, Dan Goldstein, Wolfgang Hell, Ralph Hertwig, Peter Juslin, Jürgen Locher, and Manfred Wettler for many useful comments on an earlier version of the article and Ferdinand Österreicher for help in getting the project started.

Correspondence concerning this article should be addressed to Peter Sedlmeier, Department of Psychology, Chemnitz University of Technology, 09107 Chemnitz, Germany, or to Berna Kılınç, Department of Philosophy, Boğaziçi University, 34342 Bebek, Istanbul, Turkey. E-mail: peter.sedlmeier@phil.tu-chemnitz.de or edenber@boun.edu.tr

> (b) That an athlete won the first event in the decathlon, if he won the decathlon.
>
> (c) The two events are equally probable.

According to Tversky and Kahneman, the correct answer is (c), contrary to the answers of the majority of their participants. In Tversky and Kahneman's view, that is because the prior probability for an unspecified athlete is $1/n$ (n = number of competitors), both for winning the first event in the decathlon and for winning the whole decathlon. Therefore, they stated that the two conditional probabilities must be equal. We question their argument and this conclusion in what follows. But first, we present another example.

The same problem scheme is found in Kunda and Nisbett's (1986a) studies. They asked the participants to estimate the probability that two individuals would maintain their relative rankings in a contest over a number of competitions. For instance, in one problem, it is stated that a student named Johnny received a higher grade than Danny on one spelling test during the school term. Participants were then asked to estimate the probability that Johnny would get higher grades than Danny on average in all 20 tests during that term. (It is not clear from the information given in Kunda & Nisbett, 1986a, what exactly *on average* means. Later, we specify two different ways of *being better on average*.)

In the spelling test problem, the prospects of an aggregate event (the average outcome of the comparisons between Johnny and Danny) is to be compared with the prospects of a single event (outcome of one comparison between Johnny and Danny). Each problem Kunda and Nisbett (1986a) raised about this comparison involved both directions of prediction—from a single event to an aggregate event as described above and vice versa, that is, from the information that Johnny got higher grades than Danny on average in all 20 tests to the probability that Johnny received a higher grade than Danny on 1 spelling test. According to Kunda and Nisbett, in problems such as the above, prediction symmetry holds without exception because of a "straightforward implication" of the law of large numbers known as the *aggregation principle*.

The aggregation principle, in Kunda and Nisbett's (1986a) formulation, states that "the confidence one may have in predictions increases with the size of the sample one is predicting from and with the size of the sample one is predicting to" (p. 340). They justified their claim about the symmetry of the two kinds of prediction by combining this principle with a correlation analysis: "Since the magnitude of a correlation is independent of the direction of prediction, aggregation increases predictability equally whether it is performed on the predictor or on the predicted events" (Kunda & Nisbett, 1986a, p. 341). In their studies, they transformed participants' probability estimates into Kendall's τ, which, in turn, they transformed into Spearman's r (Kunda & Nisbett, 1986b, pp. 199–201).

Let us look more carefully at the Bayesian solution of the Tversky and Kahneman (1982) example. Let W_1 denote the event of the particular athlete winning Event 1 and W denote the event of the same athlete winning the decathlon. According to Bayes's theorem, whenever the denominators below are nonzero,

$$P(W|W_1) = \frac{P(W \,\&\, W_1)}{P(W_1)}$$

and

$$P(W_1|W) = \frac{P(W \,\&\, W_1)}{P(W)}.$$

Clearly the two conditional probabilities are equal if and only if

$$P(W_1) = P(W).$$

Tversky and Kahneman (1982) assumed that this equality holds in their example. But, this equation is not forced by the mathematics of probability theory. Can we nonetheless assume it a priori to be the case? We claim not. There are good reasons to challenge this equality. Before providing a theoretical analysis of how this assumption is unjustified, we present a case study concerning soccer games. Independently of our theoretical analysis, this study challenges the empirical adequacy of the models used in the literature to argue for prediction symmetry: The case study below illustrates that the above assumption can be empirically violated.

Soccer Games: A Case Study

We chose soccer as the domain for the case study because statistical data on soccer games are readily available, and most people in Germany (where the study was performed) are familiar with the game. Soccer is played in two halves of 45 min each. In our soccer game example, the smaller sample ($N = 1$) provides the result after the first half (lead vs. no lead), and the larger sample provides the result after two halves ($N = 2$), that is, after the whole game (win vs. no win). Performance of a team, couched in this language, can be predicted on the basis of information, again described in these terms, specifying halftime or the final game results.

Study

The following problem was presented in written form to 64 University of Paderborn, Germany, students in an introductory psychology course:[1]

> Which of the following events is more likely in soccer games? (check one)
>
> __ (a) The team that leads at halftime wins the game.
>
> __ (b) The team that wins the game was in the lead at halftime.
>
> __ (c) No difference.
>
> Please comment on your solution:

Instead of being asked for the most likely event (probability instruction), 58 additional students were asked which event "occurs most often" in soccer games (frequency instruction). The overwhelming majority in both groups (81% and 83%) chose the third alternative corresponding to symmetry. Our participants had apparently found the solutions suggested by Tversky and Kahneman (1982) and Kunda and Nisbett (1986a). But, were they justified? Fortunately, we are in a position to examine this issue empirically.

Real Data

The two conditional probabilities relevant to the soccer problem are (a) that of winning the game if one has a halftime lead,

---

[1] Original versions were in German.

P(WinEnd|WinMid), and (b) that of having had a lead at halftime if one has won the game, *P*(WinMid|WinEnd). These probabilities can be estimated from the actual relative frequencies of these events. We analyzed all 308 soccer games played in the world championships between 1930 and 1978 (Huba, 1978).[2] In 243 of these games, one team won, that is, there was no tie. Of these 243 games, the winning team led at halftime in 150 games. So, *P*(WinMid|WinEnd) = 150/243 = .62. Conversely, in 197 of all the games, one team led at halftime, whereas in the other 111 games there was a tie at halftime. Of the 197 games, 153 ended with a win for the team that had the halftime lead. So, *P*(WinEnd|WinMid) = 153/197 = .78. In the domain of soccer games, it seems to be more likely that a team will win the game if it leads at halftime than that a team led at halftime if it won the game, with the difference in conditional probabilities being p = .16.

To collect more evidence, we also analyzed the results of all soccer games played during the 1994–1995 season in the German Bundesliga (national soccer league). The data were taken from the German national daily newspaper *Süddeutsche Zeitung*, which regularly reports end results as well as halftime scores. In 125 of the 200 games in which there was a winner (i.e., there was no tie), the winner led at halftime: *P*(WinMid|WinEnd) = 125/200 = .63. Of the 165 games in which one team had a lead at halftime (no tie), 127 were won by that team: *P*(WinEnd|WinMid) = 127/165 = .78. The world championship and Bundesliga analyses yield remarkably similar estimates of the two conditional probabilities. Unless one holds an extreme subjectivist interpretation of probability, one should use a frequency analysis such as ours as a corrective to one's estimates of probability. That is, under almost all interpretations of probability, these results should be taken to corroborate our claim that symmetry in prediction may fail.

Discussion

It is interesting to note that our study on soccer games could have been used to demonstrate that people understand Kahneman and Tversky's (1982) logic of symmetry or that they apply Kunda and Nisbett's (1986a, 1986b) aggregation principle, that is, that they behave "rationally."[3] However, although they would conform to a normative model, the judgments of our participants did not conform to reality. Instead of implicating the rationality of our participants, we believe this situation is because our participants did not thoroughly understand the task. Our participants' written comments suggest that many of them did not realize that outcomes described in alternatives (a) and (b) differed from each other: Of the 40 participants who commented on their solution, 21 explicitly stated that they took the two alternatives to mean the same thing. Others who chose the third alternative may have made the same interpretation and simply not commented on their solutions.[4]

Participants' performance aside, the purpose of this case study is to support our claim that symmetry in prediction does not necessarily hold and therefore should not be used as a normative criterion. This claim, hitherto illustrated solely by examples, is justified below with a formal analysis.

A Mathematical Model

The participants' task in the above mentioned studies was to compare conditional probabilities, *P*(single event|aggregate event) and *P*(aggregate event|single event). How are these probabilities determined? The probability of the single event (hereafter "single") refers to the result of an ordinal comparison of single performance values. Examples of such ordinal comparisons are whether a soccer team has scored more goals than the other at halftime, whether a competitor has more points in one event of the decathlon than his or her cocompetitors, or whether Johnny has a higher score than Danny in one spelling test. The probability of an aggregate event (hereafter "aggregate") refers to the result of an ordinal comparison of aggregate values. In the soccer game example, the comparison is between the number of goals the two teams score across the two halves of the game; in the decathlon example, the aggregate refers to the comparison of the total number of points the competitors accumulate across the 10 events of the decathlon; in the case of the spelling test, however, there are different things the aggregate ("average ranking," Kunda & Nisbett, 1986b, p. 204) may refer to. Let two competitors (e.g., Johnny and Danny) be A and B. A can be said to "win on average" (e.g., Johnny has a higher grade than Danny "on average") according to at least two different criteria. First, if the number of single wins for A is higher than the number of single wins for B, one could declare A to have won on average. This kind of success will be termed *win*. Second, as mentioned in the decathlon and soccer game examples, one can compare the sum of the scores of the two competitors: If A's final score is higher than B's final score, A can again be said to have won on average. This kind of winning will be termed *win**. For the sake of brevity, we discuss only *winning* in detail but occasionally refer to *winning** as well.[5]

We base our analysis of prediction symmetry on the following scenario. Suppose that A and B (e.g., two students) "play" (e.g., obtain a test score) a "game" (e.g., a school term) in N stages (e.g., a series of N spelling tests). On each stage i, the scores of the players are distributed according to the random variables X_i and Y_i,

---

[2] Overtime and penalty kicks (which are resorted to when overtime ends in a draw) were not included in the analysis because they are unique to the final rounds of world championships and other special tournaments.

[3] If we followed the procedure proposed by Kunda and Nisbett (1986a, 1986b), we would correlate halftime results with game results and use the coefficient as an interchangeable indicator of how well the aggregate can be predicted from the single outcome and vice versa.

[4] To get a more accurate estimate of how many participants gave the correct answer, that is, the answer consistent with our analysis of real data sets, one can exclude those who selected the third choice ("no difference"). In the probability group, only 50% (6 out of 12) of those who chose (a) or (b) found the correct answer (a)—a result that conforms with chance performance—whereas in the frequency group, 80% (8 out of 10) chose the correct answer. This is consistent with results obtained by Fiedler (1988) and Gigerenzer and Hoffrage (1995) that showed that people are far better at solving statistical problems if numerical information is presented in terms of frequencies instead of probabilities. This distinction in presentational formats has also proven useful in statistical training (Sedlmeier, 1999; Sedlmeier & Gigerenzer, 2001). A study in which the soccer game problem is explained to each participant individually to block the inference that (a) and (b) are identical might lead to a higher proportion of correct answers, especially if problems are phrased in terms of frequencies.

[5] We also analyzed a third criterion of winning: A wins the game if A wins more than $N/2$ stages in a total of N stages. For this criterion, the basic results reported below for *win* and *win** hold as well. Details of the calculations for the third winning criterion can be obtained by writing to

respectively, with $i = 1, \ldots, N$. Thus, the outcome of the *i*th stage of the game is one of the following mutually exclusive alternatives:

1. A wins if $X_i > Y_i$.
2. B wins if $X_i < Y_i$.
3. A and B tie if $X_i = Y_i$.

Assume for the sake of simplicity that $\{X_i\}$ and $\{Y_i\}$ are each independently and identically distributed and that $\{X_i\}$ is independent of $\{Y_i\}$. In that case, there are three fixed probabilities (independent of stage) for the outcomes 1, 2, and 3 above, to which we refer, respectively, as $P_\text{a} = P(X_i > Y_i)$, $P_\text{b} = P(X_i < Y_i)$, and $P_\text{t} = P(X_i = Y_i)$. Recall that the *winning* criterion is: A *wins* the game if and only if A wins more stages than B. We refer to the probability of this event as P_A.

In what follows, we show that the probability of winning in one stage ($N = 1$) is not generally equal to the probability of winning in more than one stage ($N > 1$). This is tantamount to showing that there is no symmetry in prediction. In other words, we demonstrate that the conditional probabilities that (a) A won the first game given that he or she won the multistage game and (b) A won the multistage game given that he or she won the first game will in general not be equal. We start with a simple case, comparing the result of a one-stage game to that of a two-stage game.

A Special Case

Let us illustrate the above formalism by the following coin-tossing game. In each stage of the game, two players, A and B, each flip a coin. A head corresponds to 1 point and a tail corresponds to 0 points for the player who flipped the coin. The probabilities of heads and tails are p and $1 - p$, respectively. Let us denote the number of points (0 or 1) that Players A and B obtain at the *i*th stage of the game with the random variables X_i and Y_i, respectively.

At each stage, A wins, loses, or ties with B. Assuming that X_i and Y_i are independent, then $P_\text{a} = p(1 - p)$, $P_\text{b} = p(1 - p)$, and $P_\text{t} = 1 - 2p(1 - p)$. Let us now compute the probability that A wins a two-stage game according to the *winning* criterion. According to this criterion, A will *win* either if A wins in one stage and there is a tie in the other or if A wins both stages. So,

$$P_\text{A} = 2P_\text{a}P_\text{t} + P_\text{a}^2 = p(1 - p)[2 - 3p(1 - p)].$$

Compare this result to the result in a one stage game, $P_\text{a} = p(1 - p)$. It can easily be shown that the probability of A winning in one stage (single) differs from the probability of A winning the whole game (aggregate), unless $p = 0$ or $p = 1$. If, however, P(single) and P(aggregate) differ, the conditional probabilities also differ, that is, $P(\text{single}|\text{aggregate}) \neq P(\text{aggregate}|\text{single})$. Now, we formulate this idea more generally.

General Case

The probability that Player A *wins* after N stages can be specified as follows (see the Appendix for details):

$$P_\text{A} = \sum_{k=1}^{N} \sum_{j=0}^{\min(k-1,N-k)} \binom{N}{k,\, j,\, N-k-j} P_\text{a}^k\, P_\text{b}^j\, P_\text{t}^{N-k-j}. \qquad (1)$$

The specific values of $P_\text{a} = P(X_i > Y_i)$, $P_\text{b} = P(X_i < Y_i)$, and $P_\text{t} = P(X_i = Y_i)$ to be inserted in this formula depend on the random variables X_i and Y_i.

The Impact of Ties

Could it be that prediction asymmetry is due to the existence of ties? We investigate this question by analyzing the relation between P_t and $P_\text{T} = P(\text{tie at the end of the game})$ for the *winning* criterion. When the possibility of a tie at one stage is ruled out, that is, when $P_\text{t} = 0$, the possibility of a tie at the end of the game is not ruled out unless N is odd. If N is even, then a tie holds just in case A and B each win exactly $N/2$ stages of the game. So, we have[6]

$$\text{If } P_\text{t} = 0, \text{ then } P_\text{T} = \begin{cases} 0, \text{ if } N \text{ is odd} \\ \binom{N}{N/2} P_\text{a}^{N/2} P_\text{b}^{N/2}, \text{ if } N \text{ is even} \end{cases}.$$

Conversely, if $P_\text{T} = 0$, then either N is odd or both $P_\text{t} = 0$ and either $P_\text{a} = 0$ or $P_\text{b} = 0$ (see the Appendix for details). That is, whenever N is even and either $P_\text{t} \neq 0$ or both $P_\text{a} \neq 0$ and $P_\text{b} \neq 0$, it follows that $P_\text{T} \neq 0$.

Consider the situation in which $P_\text{a} = P_\text{b} = 1/2$. This means that $P_\text{t} = 0$. If N is even, the above formula yields the probability of a tie at the end of the game to be

$$\binom{N}{N/2} \frac{1}{2^N}.$$

Under these circumstances, clearly $P_\text{a} \neq P_\text{A}$. In this case, $P_\text{T} \neq 0$, and one may wonder if prediction asymmetry disappears in games in which $P_\text{T} = 0$. It can be shown that even when both $P_\text{t} = 0$ and $P_\text{T} = 0$, we cannot conclude that $P_\text{a} = P_\text{A}$. An example suffices to show this: Suppose that we estimate the chances of Player A winning a stage to be less than those of Player B—for example, that $P_\text{a} = 1/3$ and $P_\text{b} = 2/3$—and that there are five stages in a game, that is, $N = 5$. Under these circumstances, both $P_\text{t} = 0$ and $P_\text{T} = 0$. It can be shown that in this situation, $P_\text{A} = .21$. (To obtain this figure up to two significant digits, just substitute $P_\text{t} = 0$, $P_\text{a} = 1/3$, $P_\text{b} = 2/3$, and $N = 5$ into Equation 1). Thus, even when $P_\text{t} = 0$ and $P_\text{T} = 0$, prediction symmetry may fail. Our examples here demonstrate that prediction symmetry holds only as a special case in the domain of games we model.

Further Specifications

Although the above findings suffice to refute as too simplistic the received model used in the relevant literature, we continue with some further elaboration of our model to indicate the ways in which we can capture the situations encountered in everyday prediction tasks. To do so, we analyze Equation 1 by assuming some specific distributions for the random variables X_i and Y_i. More precisely, we assume X_i and Y_i, for each i, to be binomial

---

the authors. An analogous winning criterion that applies, for instance, to volleyball is that the first player who wins more than N stages "wins" the game.

[6] This result also follows from Equations A1 and A3 in the Appendix when the value of P_t is taken as 0.

random variables with distributions $B(n, p)$ and $B(n, q)$, respectively. In this way, we can represent some further specifics of a "game," such as the maximum range or score, n, at each stage, and the respective "skills," p and q, of the players.

In what follows, we examine the factors that figure in the above model with a view to specifying the most important parameters responsible for prediction asymmetry. For that purpose, we pay special attention to sample size (N), range (n), and skills of the competitors (p and q).

Sample Size (Number of Stages)

We begin by comparing the different values of the probability that A *wins* a multistage game, when both A and B have the same skill represented by $p = q = .50$, while the number of stages N is varied. Because in the case of binomial distributions these probabilities also depend on the range, n, we computed P(A *wins* after N stages with each stage ranging up to n) as a function of n. Figure 1 shows the values of this probability for some selected numerical values of N, plotted against range n. Prediction symmetry would be obtained if the probability that A *wins* were independent of the number of stages of the game. That would be represented in Figure 1 by the coincidence of all the curves. This is clearly not the case. The probability that A *wins* does generally depend on sample size N.

Note the way Figure 1 captures the coin-tossing game discussed above. Recall that in this example, the range n is 1 (at each stage one can win up to 1 point), and the sample sizes N that are compared are 1 and 2. Thus, looking at the curves for $N = 1$ and $N = 2$, we can locate the values of P_a and P_A at $n = 1$, that is, at the y-axis. In conformity with our previous calculations, Figure 1 shows that $P_a = P(\text{single}) = .25$ and $P_A = P(\text{aggregate}) = .3125$. Because, according to Bayes's theorem, $P(\text{single}|\text{aggregate}) = [P(\text{single})/P(\text{aggregate})]P(\text{aggregate}|\text{single})$, we have for the coin-tossing game, $P(\text{single}|\text{aggregate}) = 0.80P(\text{aggregate}|\text{single})$. That means that in this case, one can predict more confidently from the smaller to the larger sample than the other way around.

As another example, let us consider a 10-stage game ($N = 10$) with a range of 50 and investigate the task of predicting from or to a part of the game that consists of $N = 3$ stages (see arrows in Figure 1). This time, our curves indicate that one can be more confident in the prediction from the larger to the smaller sample. Because $P(\text{smaller}) = .449$ and $P(\text{larger}) = .425$, we find for this case that $P(\text{smaller}|\text{larger}) = 1.06P(\text{larger}|\text{smaller})$. If, again with a range of 50, a 10-stage game was compared to a 1-stage game, we would obtain $P(\text{single}|\text{aggregate}) = 1.08P(\text{single}|\text{aggregate})$.

In general, the ratios $P(\text{single})/P(\text{aggregate})$ or $P(\text{smaller})/P(\text{larger})$ determine the prediction asymmetry: If this ratio is less than 1, as in the coin-tossing example above, then the smaller sample is a better indicator of overall performance. If this ratio is larger than 1, then the larger sample is a better indicator of partial performance. And finally, when this ratio equals 1, we have prediction symmetry—the intuition expressed by our participants

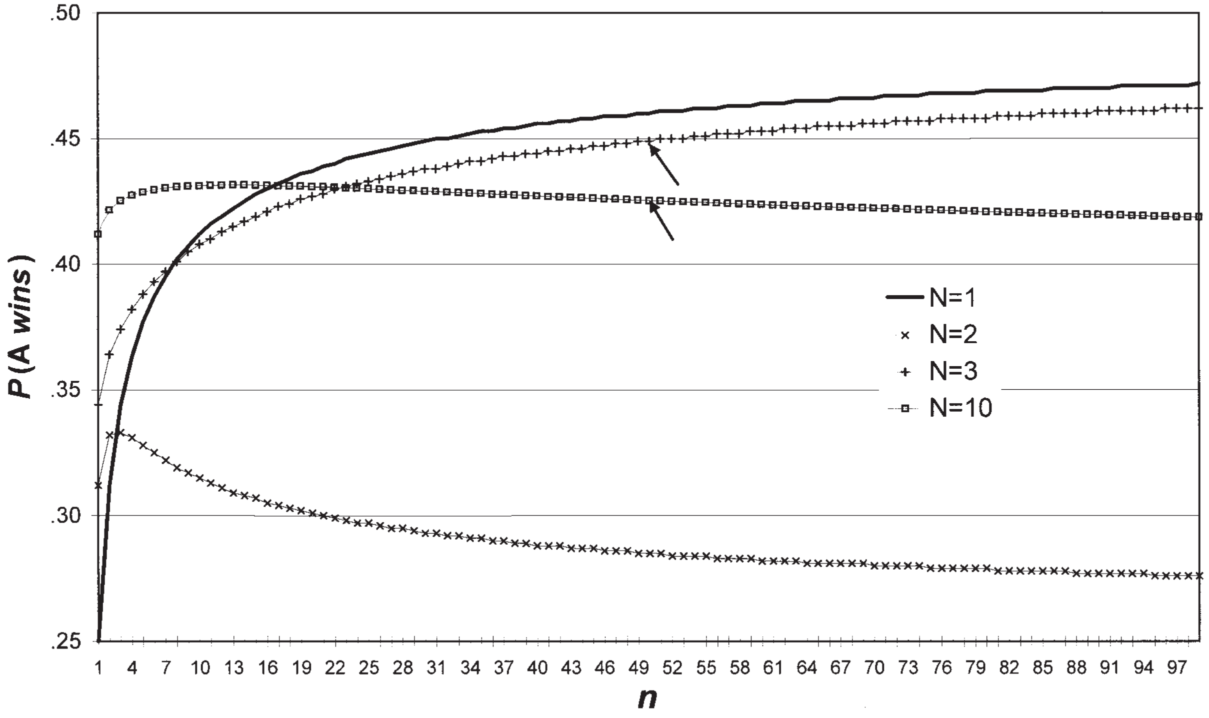

Figure 1. Probability of A winning over B according to the *winning* criterion. For $N = 1$, this probability is P_a, and for values of N larger than 1, this corresponds to what is denoted by P_A in the text. At each stage, A's and B's possible scores follow the binomial distribution $B(n, .50)$. The variables are range (n = 1–99, in steps of 1) and sample size (Ns = 1, 2, 3, and 10). The arrows indicate the probabilities of A winning over B for a range of $n = 50$ and two different numbers of stages: $N = 3$ (upper arrow) and $N = 10$ (lower arrow). See the text for further explanation.

in the soccer study and in former research. This suggests that another way we can visualize skewed symmetry is by means of a graph of *P*(single)/*P*(aggregate). In Figure 2, we plotted the ratios *P*(single)/*P*(aggregate) for the case of $p = q = .50$ and the range *n*, assuming five different values, against the sample size *N*, varying from $N = 1$ to $N = 40$ (*x*-axis in Figure 2). Prediction symmetry obtains only when the graphs cross the horizontal line represented by *y*-axis = 1. We can display here from another angle some of the results discussed above: With a range of $n = 1$ and sample size of $N = 2$, the ratio *P*(single)/*P*(aggregate) is 0.80, and with a range of $n = 50$ and a sample size of $N = 10$, this ratio is 1.08 (see arrows in Figure 2). Figure 2 shows once again our finding that in this special case prediction symmetry is an exception, although it appears that there is approximate prediction symmetry for ranges larger than 25 and sample sizes larger than 15.

Difference in Skill

That two competitors have equal skills is not a very plausible assumption for a wide range of comparisons found in everyday life. Does the asymmetry in predictions found for the case in which $p = q$ (see Figures 1 and 2) also hold in the more realistic case when *p* differs from *q*? We consider the question for the case in which Player A does slightly worse than Player B, that is, for $p = .45$ and $q = .50$. Our results (see Figure 3) confirm our previous conclusion: Asymmetry prevails also on the assumption of different skills. In point of fact, difference in skills amplifies asymmetry to such an extent that we were not able to squeeze in the results for values of *n* larger than 10 in the same graph. For instance, when $N = 40$, we obtain ratios of 12.8 and 82.4 for the range values of 25 and 50, respectively. Recall that this means

$$P(\text{aggregate}|\text{single}) < P(\text{single}|\text{aggregate}).$$

That is, for these ranges, the prediction from the aggregate to the single is vastly more reliable than the reverse prediction.

Range

Situations in everyday life that we represent as games involve a wide diversity of ranges. In the coin-tossing game the range is 1, whereas in a spelling test the range may be on the order of 20. What is the impact of the magnitude of range on the asymmetry in predictions? Figures 1–3 show that the range can be crucial in the determination of the direction of asymmetry, that is, whether the prediction is better from the smaller to the larger sample or the reverse. Varying the range, we can change the direction of asymmetry. When the curves in Figure 1 cross, there is a reversal of the direction of asymmetry. It thus follows that, at least in setups similar to ours, range must be taken as another decisive factor in the specification of the normative model.

The Win* *Criterion*

The above calculations were all based on the *winning* criterion, according to which Player A is designated to be the winner when A wins more stages of a multistage game than Player B. Several comparisons encountered in daily life are based on another standard of winning that we introduced above as the *winning** crite-

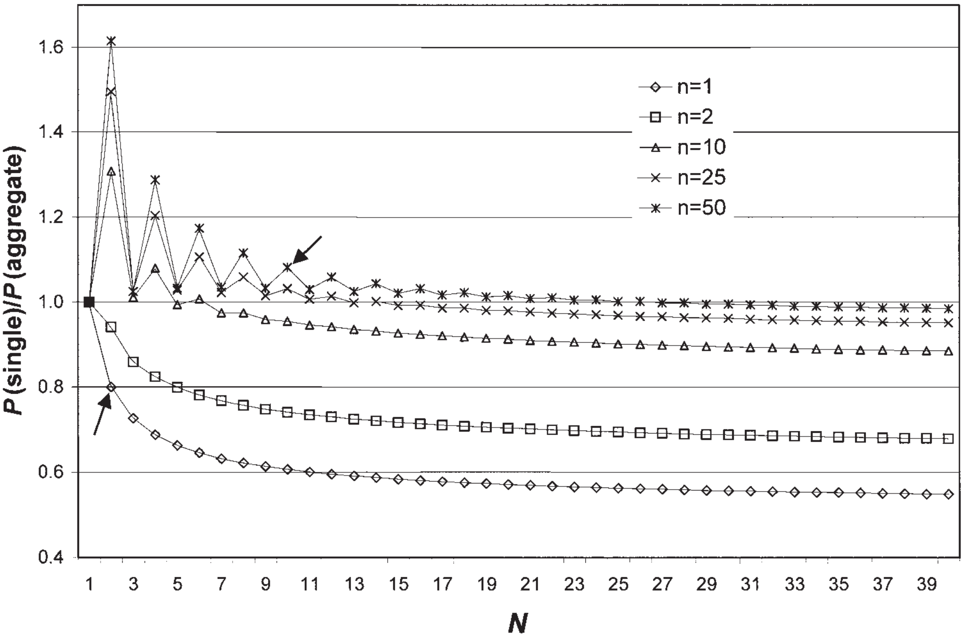

Figure 2. Prediction asymmetry from another perspective, according to the *winning* criterion. Sample size is varied on the *x*-axis from $N = 1$ to $N = 40$, in steps of 1, and results are shown for ranges of *n*s = 1, 2, 10, 25, and 50. A's and B's possible scores each follow the binomial distribution $B(n, .50)$. The ratio *P*(single)/*P*(aggregate) is larger than 1 if and only if the prediction is better from the aggregate result to the single case. The arrows indicate this ratio for a range of $n = 1$ and a sample size of $N = 2$ (left arrow) and for a range of $n = 50$ and a sample size of $N = 10$ (right arrow). See the text for further explanation.

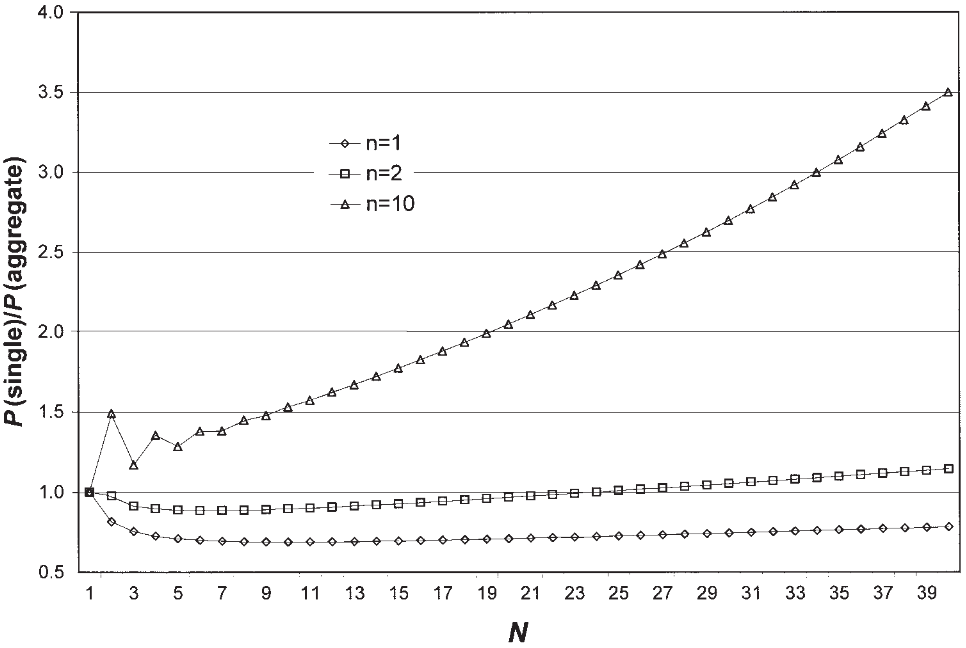

Figure 3. Prediction asymmetry, according to the *winning* criterion, when Players A and B have different skills. *P*(aggregate) denotes *P*(A *wins* after *N* stages). A's possible scores at each stage follow the binomial distribution $B(n, .45)$, whereas B's possible scores follow $B(n, .50)$. Results for ranges of *n*s = 1, 2, and 10 are shown.

rion. In our model, the probability that Player A wins according to this criterion, to be denoted by P^*_A, is

$$P^*_A = P\left[\sum_{i=1}^{N} X_i > \sum_{i=1}^{N} Y_i\right].$$

It is difficult to break up this formula analytically into components involving P_a, P_b, and P_t like Equation 1 for the *winning* criterion. To facilitate calculations, we consider again the case in which X_i and Y_i are binomial random variables $B(n, p)$ and $B(n, q)$, respectively (see the Appendix for details). Hence, we expect the same parameters of interest to have an impact on our results. Our calculations indicate that the basic conclusion we have reached above holds also for the *winning** criterion: Asymmetry in predictions is the rule and not the exception, and the exact discrepancy depends on the same parameters of interest, namely *n*, *p*, *q*, and *N*. Figure 4 illustrates special cases, with ranges of $n = 1$ and $n = 2$ and the number of stages *N* varied as before. The ratio *P*(single)/*P*(aggregate) assumes different values depending on the skills of the two players also; so, we have included different trajectories, corresponding to (a) equal skills, $p = q = .50$, and (b) different skills, $p = .45$ and $q = .50$. It can be seen that for almost all stages, the graphs deviate considerably from the unit value 1, which would correspond to prediction symmetry.

Elimination of Ties

We have already shown that ties cannot be easily discounted at the aggregate level, even if it is possible to encounter situations that exclude them in a single stage. But, perhaps we can completely eliminate ties by fiat. Here is one way to do so in the case of two players: Count ties randomly as success or failure for a given player. Thus, one can stipulate that if there is a tie in one stage, one tosses a fair coin to decide whether that stage is to count as a win for Player A. Similarly, at the end of the game, if there is a tie, throw a fair coin to decide whether Player A is to be declared the winner.[7] Under this stipulation, prediction symmetry would result when players have equal skills because probabilities will indeed be equally divided. However, it is unclear to what extent real-life situations fall under this new model. One can imagine soccer games conducted under this new winning criterion, say by penalty kicks (by equally good players to equally good goalkeepers) at the end of the halftime and then again at the end of the game, but in many other situations, even a stretch of imagination does not suffice to apply this new model. Moreover, if the players have different skills, this stipulation does not save prediction symmetry. Figure 5 shows that for the *winning* criterion, already with slightly different skills, that is when $p = .45$ and $q = .50$, the elimination of ties by random division still leads to asymmetry, which becomes more pronounced when the range gets larger.

Application

The basic finding in the above analysis is that asymmetry reigns in predictions involving different sample sizes. Furthermore, the

---

[7] This is tantamount to introducing a new winning criterion, say, *win'*, which generates new formulas P'_a and P'_A out of the probabilities for the *winning* criterion as follows: $P'_a = P_a + 0.5\ P_t$ and $P'_A = P_A + 0.5\ P_T$. Analogous extensions hold for the *winning** criterion as well.

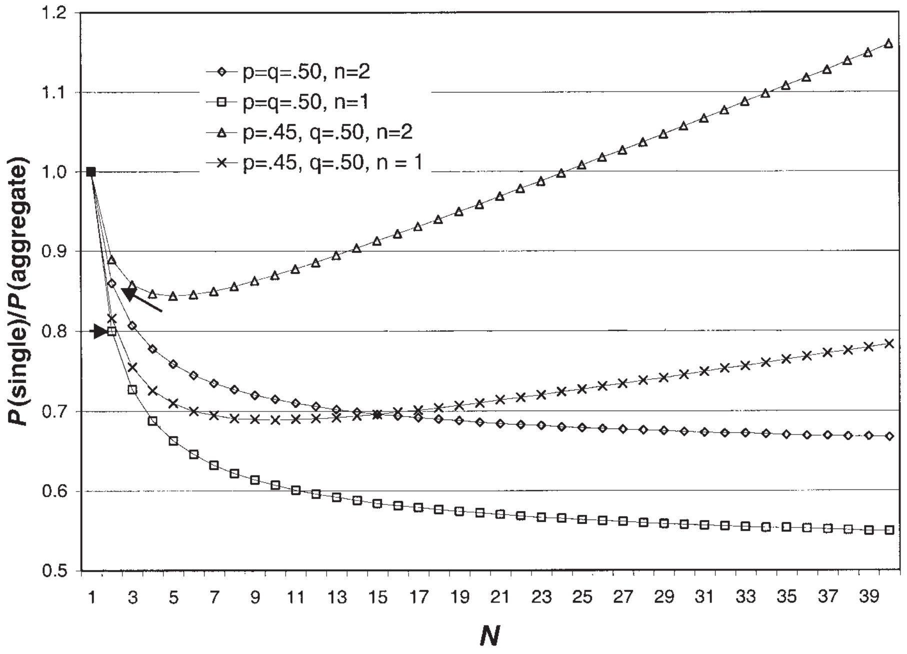

Figure 4. Prediction asymmetry for Players A and B according to the *winning** criterion. A's and B's possible scores follow binomial distributions $B(n, p)$ and $B(n, q)$, respectively. Results are shown for ranges of $n = 1$ and $n = 2$ and $p = q = .50$ as well as $p = .45$, $q = .50$. The two arrows refer to asymmetries that could be expected in soccer games with up to one (lower arrow) and up to two (upper arrow) goals per halftime. See the text for further explanation.

amount and the direction of the asymmetry follow a complex pattern that depends, among others, on the skills of the players, on the range of comparison at each stage, and on the number of stages. How do the results of this formal analysis relate to the decathlon, spelling test, and soccer game problems discussed earlier? In decathlon and soccer games, the winning criterion is what we have called the *winning** criterion, whereas in the spelling test, what is at stake is the *winning* criterion. Let us apply our separate findings for these criteria above to these competition domains.

Decathlon, Spelling Test, and Soccer

One might argue that in Tversky and Kahneman's (1982) decathlon problem, prediction symmetry holds approximately because in a single decathlon event the range may be up to 800. But usually, participants in a major decathlon competition can all be expected to reach a minimum of, say, 600–700 points in all disciplines, and therefore, the meaningful range can be expected to be about 100 points. Our analysis shows that the differences between conditional probabilities become quite small for such ranges, at least if the competitors are equally good. For instance, for a range of $n = 100$ and $p = q = .50$, according to the *winning** criterion, $P(\text{single})/P(\text{aggregate}) = 0.96$, a figure fairly close to 1. However, actual competitors almost certainly differ in their skills. If, for example, $n = 100$, $p = .50$, and $q = .60$, that is, Player B is better than Player A, Player A's chance to *win** 1 stage is $P_a = .068$, whereas after 10 stages, P^*_A is basically 0. In this case, a prediction from the total performance in a decathlon to a single event thereof is bound to be far more successful than the reverse prediction.[8]

In the spelling test example, the point range such as that in Kunda and Nisbett's (1986a) problem can be assumed to be on the order of 10–50. For such a range, there is a nonnegligible difference between the probability for being better in 1 spelling test ($N = 1$) and that for being better on average in 20 spelling tests ($N = 20$), especially if students differ even minimally in skills (see Figures 2 and 3). This also implies considerable asymmetry in the corresponding predictions.

Finally, the soccer game problem can be modeled by the *winning** criterion for $N = 2$ by fixing appropriate values for the range n and the skills of the teams, p and q. Of course, the latter values cannot be fixed a priori without a statistical analysis of the performance of teams and scores on halftimes. Yet, to illustrate the structure of our model, we took $n = 1$, $n = 2$, and $p = q = .50$ and found that $P^*(\text{WinMid})/P^*(\text{WinEnd})$ equals 0.80 and 0.86 for $n =$

---

[8] The applicability of our model to the decathlon problem may be questioned in view of the implicit assumption that each athlete has the same skill on each discipline. The decathlon can be modeled better by taking each X_i to have a $B(n, p_i)$ distribution, where p_i indicates A's skill on the ith discipline, say on the long jump. This would make for a more complicated model, the analysis of which we leave out in this article for want of space.

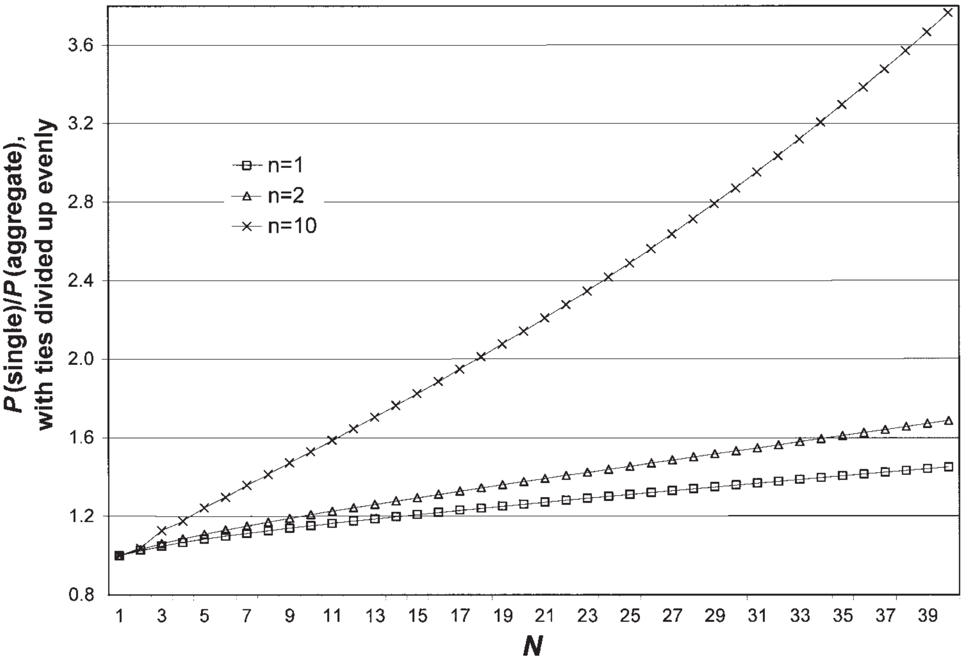

Figure 5. Elimination of ties when A and B have different skills. A's and B's possible scores each follow binomial distributions $B(n, .45)$ and $B(n, .50)$, respectively. Results are shown for ranges of ns = 1, 2, and 10, according to the *winning* criterion.

1 and $n = 2$, respectively (see arrows in Figure 4).[9] These quotients correspond to $P(\text{WinMid}|\text{WinEnd})/P(\text{WinEnd}|\text{WinMid})$, the empirical values of which were found to be .79 and .81 (.62/.78 and .63/.78, see above). The close match between our model and the empirical results in soccer games corroborates our approach. The residual difference may be due to other factors, such as the overall ranking, tactics, the coach, or the mood of the players. One should also take into account some sort of endurance factor to model that some teams that are behind at halftime tend to get discouraged, whereas others may work even harder. These considerations call for even more complex models to analyze games like soccer.

Limits of Analysis

Our model was spelled out in the vocabulary of games. That is only a metaphorical device and does not mean that its applications are restricted to games only. Many ordinary life situations involving comparison of outcomes on the average can be fit into this framework, once we extend the meanings of *game*, *stage*, and *skill*. For instance, a manager might want to predict which of two candidates for a job will perform better on the average from having them work for one day. Or, he or she might want to predict which of two employees' performance will be better in the next week from their average performance during the last 4 years. If one would regard their performance on a given day as a stage in a game of many days, then our analysis would be applicable to this example as well.

Our analysis can be also extended by taking into account distributions other than the binomial ones as representative of the "performance" of "competitors." One could, for instance, imagine cases in which the number of "points" A scores in a single stage of the game is distributed in accordance with a Poisson distribution. Although the consideration of different distributions would generate or remove some parameters, it seems not too unrealistic to expect something analogous to range and skill to figure in a variety of probability distributions deemed applicable to these situations.

In our mathematical models, we assumed that the result in each stage of the game was independent of the results of other stages. This assumption is probably unwarranted in many cases. Many real-world problems deal with samples or stages that depend on the previous ones (stochastic processes). However, we do not expect our conclusion—that asymmetry in everyday predictions is the general rule—to be invalidated by dependence, temporal development, or noisy distributions. Such factors render complete specification of a normative model all the more difficult, or even impossible.

Conclusion

We have illustrated that numerous and sometimes insurmountable difficulties may arise when a complete mathematical specification of a normative model is sought to represent correct decision making. That is so, even when we waive concerns about the choice of a normative theory of decision making, such as a choice between Fisherian and Bayesian statistical theories. Basing predictions about one sample on a sample of different size is not the

---

[9] In soccer games with highly skilled teams, such as those who play in the German Bundesliga and in world championships, 0 to 2 goals (a range of 1 or 2) per team in one halftime are very common results.

only case in which a complete specification of the normative model poses such difficulties. Analogous problems arise, for instance, in the probability problems discussed by Bar-Hillel and Falk (1982), who found that models involving conditional probabilities may remain underspecified unless one takes into account the way information is obtained. Underspecification calls for procedurally detailed and possibly work-intensive clarification of the ambiguities in a coarse-grained description.

The specification of a normative model is important because in psychological research deviance from such models is often seen as an index of irrationality (see Gigerenzer, 1991; Hilton, 1995; Lopes & Oden, 1991). Such an assessment, however, tends to be too hasty if one finds that the problem is underdescribed to the participants. We are not as skeptical as Lopes (1991), who concluded that "it is only in the abstract world of what have been called 'urns and balls' or 'box models' that there are clear-cut answers to probability problems" (p. 77). We always need to model to render tractable a problem encountered in real life, and modeling always involves some abstraction. Ideally, for each abstractive choice, there may correspond some normative model prescribing decision making for that way of abstraction. There is no guarantee, however, that each abstraction yields a unique model in this way—that is, no guarantee that a particular way of abstracting does not underdetermine the solution. Our worry is that when we abstract less, that is, when a problem situation is specified in its everyday complexity and analyzed with due rigor, the resulting normative model may be beyond the pale of everyday cognition. The moral we draw from results like ours is not so contentious but important: Lest we expect omniscience in everyday decision making, greater care and precision are needed in the analysis of a situation in terms of a normative model that is thereafter used to pass a verdict on human rationality.

References

Bar-Hillel, M., & Falk, R. (1982). Some teasers concerning conditional probabilities. *Cognition, 11,* 109–122.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *American Journal of Psychology, 96,* 85–94.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences, 4,* 317–331.

Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research, 50,* 123–129.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." *European Review of Social Psychology, 2,* 83–115.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102,* 684–704.

Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989). *The empire of chance: How probability changed science and everyday life*. Cambridge, England: Cambridge University Press.

Hilton, D. J. (1995). The social context of reasoning: Conversational inference and rational judgment. *Psychological Bulletin, 118,* 248–271.

Huba, K.-H. (Ed.). (1978). *1930–1978. Die Geschichte der Fuβball-WM* [*1930–1978. The history of the soccer world-championships*]. München, Germany: Copress-Verlag.

Kunda, Z., & Nisbett, R. E. (1986a). Prediction and the partial understanding of the law of large numbers. *Journal of Experimental Social Psychology, 22,* 339–354.

Kunda, Z., & Nisbett, R. E. (1986b). The psychometrics of everyday life. *Cognitive Psychology, 18,* 195–224.

Lopes, L. L. (1991). The rhetoric of irrationality. *Theory & Psychology, 1,* 65–82.

Lopes, L. L., & Oden, G. D. (1991). The rationality of intelligence. In E. Eels & T. Maruszewski (Eds.), *Poznan studies in the philosophy of the sciences and the humanities* (Vol. 21, pp. 225–249). Amsterdam: Rodopi.

MacDonald, R. R. (1986). Credible conceptions and implausible probabilities. *British Journal of Mathematical Psychology, 39,* 15–27.

Sedlmeier, P. (1999). *Improving statistical reasoning: Theoretical models and practical implications*. Mahwah, NJ: Erlbaum.

Sedlmeier, P., & Gigerenzer, G. (2001). Teaching Bayesian reasoning in less than two hours. *Journal of Experimental Psychology*: *General, 130,* 380–400.

Tversky, A., & Kahneman, D. (1982). Causal schemas in judgments under uncertainty. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 117–128). Cambridge, England: Cambridge University Press.

(*Appendix follows*)

Appendix

Derivation of Formulas

Probability for Win

The probability that A wins more stages than B in a game consisting of N stages can be calculated as follows:

$$\begin{aligned}
P_A &= P(\text{A wins more stages than B})\\
&= \sum_{k=1}^{N} P(\text{A wins } k \text{ stages and B wins}\\
&\qquad \text{less than } k \text{ stages in a total of } N \text{ stages})\\
&= \sum_{k=1}^{N} \sum_{j=0}^{\min(k-1,N-k)} P(\text{A wins } k \text{ stages, B wins } j \text{ stages, and}\\
&\qquad N-k-j \text{ stages results in a draw})\\
&= \sum_{k=1}^{N} \sum_{j=0}^{\min(k-1,N-k)} \binom{N}{k,\, j,\, N-k-j} P_a^k P_b^j P_t^{N-k-j}. \qquad \text{(A1)}
\end{aligned}$$

The upper limit for the second summation is obtained by noting that j has to satisfy both $0 \leq N - j - k$ and $j \leq k-1$. So, $j \leq \min(k-1, N-k)$. In the last step, we made use of the assumption that $\{X_i\}$ and $\{Y_i\}$ are each a sequence of independently and identically distributed random variables, with $\{X_i\}$ independent from $\{Y_i\}$. By replacing the subscripts a and b, we can obtain an analogous formula for P_B, the probability that Player B wins the game after N stages. Because of the situation they model, in these formulas, we take $0^0 = 1$; so that when, for example, $P_t = 0$, $P_t^0 = 1$. In other words, because $P_t = 0$ means that the probability of a tie at a single stage is 0, it follows that $P_t^0 = P$ (no tie at a single stage) $= 1$, and all occurrences of P_t in Equation A2 can be simply deleted.

The probability for a tie in a single stage is

$$P_t = 1 - P_a - P_b. \qquad \text{(A2)}$$

The probability for a tie at the end of the game, P_T, is found by noting either that

$$P_T = 1 - P_A - P_B \qquad \text{(A3)}$$

or that for a tie to occur at the end of N stages, N has to be even and both A and B score the same number of wins at the end of N stages. So,

$$P_T = \begin{cases} 0, \text{ if } N \text{ is odd} \\ \displaystyle\sum_{k=0}^{N/2} P(\text{both A and B win } k \text{ stages and} \\ \qquad N-2k \text{ stages result in a tie}), \text{ if } N \text{ is even} \end{cases}. \qquad \text{(A4)}$$

Again, using the independence assumption, we unpack this formula as

$$P_T = \begin{cases} 0, \text{ if } N \text{ is odd} \\ \displaystyle\sum_{k=0}^{N/2} \binom{N}{k,\, k,\, N-2k} P_a^k P_b^k P_t^{N-2k}, \text{ if } N \text{ is even} \end{cases}. \qquad \text{(A5)}$$

Note that if N is even, then $P_T = 0$ implies that each of the summands in the above summation is 0. If each summand is 0, then in particular when $k = 0$, we find $P_t^N = 0$. So, $P_t = 0$. Also, taking $k = N/2$, we get $P_a^{N/2} P_b^{N/2} = 0$, which implies that either $P_a = 0$ or $P_b = 0$, hence, the necessary conditions for $P_T = 0$ as mentioned in the text. (In these derivations, we again made use of the convention $0^0 = 1$, to capture the situations in which P_a, P_b, or P_t equals 0.)

We derive now the probabilities P_a, P_b, and P_t to be inserted into the formula for P_A in the special case where $X_i \sim B(n, p)$ and $Y_i \sim B(n, q)$ for each i. In this case,

$$\begin{aligned}
P_a &= P(X_i > Y_i)\\
&= \sum_{k=1}^{n} P(X_i = k,\, Y_i < k)\\
&= \sum_{k=1}^{n} P(X_i = k) P(Y_i < k)\\
&= \sum_{k=1}^{n} \sum_{j=0}^{k-1} P(X_i = k) P(Y_i = j)\\
&= \sum_{k=1}^{n} \sum_{j=0}^{k-1} \binom{n}{k} p^k (1-p)^{n-k} \binom{n}{j} q^j (1-q)^{n-j}\\
&= \sum_{k=1}^{n} \sum_{j=0}^{k-1} \frac{(n!)^2 p^k (1-p)^{n-k}\, q^j (1-q)^{n-j}}{k!(n-k)!j!(n-j)!}. \qquad \text{(A6)}
\end{aligned}$$

Analogously, the probability of B winning a single stage is

$$P_b = \sum_{k=1}^{n} \sum_{j=0}^{k-1} \frac{(n!)^2 q^k (1-q)^{n-k}\, p^j (1-p)^{n-j}}{k!(n-k)!j!(n-j)!}. \qquad \text{(A7)}$$

Probability for *Win**

Note that if each $X_i \sim B(n, p)$, then $\sum_{i=1}^{N} X_i \sim B(Nn, p)$. Thus for this case,

$$\begin{aligned}
P^*_A &= P\left(\sum_{i=1}^{N} X_i > \sum_{i=1}^{N} Y_i\right)\\
&= P(U > V), \text{ where } U \sim B(Nn, p) \text{ and } V \sim B(Nn, q)\\
&= \sum_{k=1}^{Nn} P(U = k \text{ and } V < k)\\
&= \sum_{k=1}^{Nn} \sum_{j=0}^{k-1} \frac{[(Nn)!]^2\, p^k (1-p)^{Nn-k}\, q^j (1-q)^{Nn-j}}{k!(Nn-k)!j!(Nn-j)!}. \qquad \text{(A8)}
\end{aligned}$$

Received September 4, 1996
Revision received December 21, 2003
Accepted December 22, 2003 ■

Correction to Sedlmeier and Kılınç (2004)

On p. 772 (*Real Data* section) of the article ‘The Hazards of Underspecified Models: The Case of Symmetry in Everyday Predictions,” by Peter Sedlmeier and Berna Kılınç (*Psychological Review*, 2004, Vol. 111, No. 3, pp. 770–780), some numbers were listed incorrectly. The correct numbers are as follows:

For soccer world championships (1930–1978), number of games in which a given team both won and led at halftime = 151, number of games in which a team won the game = 243, number of games in which a team led at halftime = 196, $P(\text{WinMid}|\text{WinEnd}) = .62$, and $\boldsymbol{P}(\text{WinEnd}|\text{WinMid}) = .77$.

For the German Bundesliga (1994–1995), number of games in which a given team both won and led at halftime = 132, number of games in which a team won the game = 223, number of games in which a team led at halftime = 174, $P(\text{WinMid}|\text{WinEnd}) = .59$, and $P(\text{WinEnd}|\text{WinMid}) = .76$. The conditional probabilities are referred to again on p. 778. There, it should read ‘These quotients correspond to $P(\text{WinMid}|\text{WinEnd})/P(\text{WinEnd/WinMid})$, the empirical values of which were found to be .81 and .78 (.62/.77 and .59/.76, see above).”