Organizational Behavior and Human Decision Processes
Vol. 85, No. 1, May, pp. 109–134, 2001
doi:10.1006/obhd.2000.2934, available online at http://www.idealibrary.com on IDEAL®

The Influence of Alternative Outcomes on Gut-Level Perceptions of Certainty

Paul D. Windschitl and Michael E. Young

University of Iowa

Recent research has demonstrated that the perceived certainty of a focal outcome depends not only on the overall amount of evidence supporting the alternatives to the focal outcome, but also on how that evidence is distributed across those alternatives (e.g., Windschitl & Wells, 1998). Three experiments replicated this alternative-outcomes effect across a variety of evidence distributions and investigated a heuristic comparison account for the effect. Participants provided gut-level certainty estimates for winning hypothetical raffles in which they and several other players held specified numbers of tickets. Results revealed that alternative-outcomes effects are not dependent on variations in the rank-order status of the focal outcome (Experiment 1) and are reliable but reduced in magnitude when the focal outcome is the least likely outcome (Experiment 2). Also, consistent with a core premise of the heuristic comparison account, evidence supporting the strongest alternative outcome was shown to play the primary role in producing alternative-outcomes effects (Experiment 3). © 2001 Academic Press

Imagine a situation for which there are four possible outcomes. According to prescriptive and descriptive analyses, people's perceptions of certainty about a focal outcome should be, and generally are, sensitive to how the evidence for that focal outcome compares to the aggregated evidence for the alternative

Preparation of this paper was supported by Grant SES 99–11245 from the National Science Foundation. We are grateful to Brad Brunick, Annette Flugstad, and Mary Jenson for their comments and help on various parts of this research.

Address correspondence and reprint requests to Paul D. Windschitl, Department of Psychology, University of Iowa, Iowa City, IA 52242. E-mail: paul-windschitl@uiowa.edu. Current address: Michael Young, Department of Psychology, Southern Illinois University, Carbondale, IL.

0749-5978/01 $35.00

outcomes.[1] However, recent studies indicate that people's perceptions of certainty about the focal outcome are also sensitive to variations in how the evidence for the alternative outcomes is distributed, even when such variations have no bearing on the objective probability of the focal outcome (Gonzalez & Frenck-Mestre, 1993; Teigen, 1988, in press; Windschitl & Wells, 1998).

The effect that these variations have on perceptions of certainty has been called the *alternative-outcomes effect* (Windschitl & Wells, 1998). In one study demonstrating the effect, participants read a scenario that described a situation in which they held 21 tickets in a raffle (Windschitl & Wells, 1998). In one version of the scenario, five other raffle players held 14, 13, 15, 12, and 13 tickets; in another version, the five other players held 52, 6, 2, 2, and 5 tickets. Although the objective probability of winning in the two scenarios was identical (23.9%), participants reading about the former distribution (21-14-13-15-12-13) expressed more certainty than participants reading about the latter distribution (21-52-6-2-2-5). In another study, participants read about a girl who hoped to pick a chocolate-chip cookie from a cookie jar. They expressed less certainty in that focal outcome when they read that the jar contained two chocolate-chip cookies and seven oatmeal cookies (2-7) than when the jar contained two chocolate-chip cookies, one oatmeal, one raisin, one butterscotch, one rum, one peanut butter, one pecan, and one sugar cookie (2-1-1-1-1-1-1-1). In related research, Teigen (1988, in press) has described the *equiprobability effect*. Demonstrations of the equiprobability effect suggest that people often feel quite optimistic about the chances of a focal event if the focal event is equiprobable to all the possible alternative events, but not when the alternatives have unequal probabilities.

There are many conceivable ways of explaining why manipulations to the distributions of alternative outcomes influence the perceived likelihood of a focal outcome. For example, Teigen (in press) has proposed a propensity interpretation to explain the equiprobability effect. According to this interpretation, the perceived likelihood of a focal outcome is based on an assessment of the presence/absence of causal factors that could produce or inhibit that outcome. In a situation in which several possible outcomes are equiprobable, a focal outcome can seem to have a good chance of occurring because factors facilitating the occurrence of the focal outcome are present and there are no factors strongly inhibiting its occurrence (Teigen, in press).

The present article focuses on a different type of explanation that can be readily applied to predicting how people would respond to a wide range of evidence distributions (one of which might be an equiprobable distribution).

[1] Although there is clear evidence that people sometimes underestimate or overestimate the total amount of evidence for alternative outcomes and that their probability estimates for a set of mutually exclusive and exhaustive outcomes do not exhibit perfect complementarity, the evidence also suggests that perceptions of a focal outcome's certainty are, in significant measure, influenced by the strength of the support for the alternatives (e.g., see Tversky & Koehler, 1994; Sanbonmatsu, Posavac, & Stasney, 1997; Windschitl, 2000).

This explanation, proposed by Windschitl and Wells (1998), assumes that pairwise comparisons between the focal outcome and individual alternative outcomes can play an important role in the perceived certainty of a focal outcome. More specifically, a comparison between the focal outcome and the strongest alternative outcome can sometimes have a dominant role in determining the perceived certainty of the focal outcome. The more this comparison favors the focal outcome (or the less it favors the most likely alternative), the greater the perceived likelihood of the focal outcome. For example, according to this explanation, participants in Windschitl and Wells's (1998) raffle-ticket study felt more certain about winning in the 21-14-13-15-12-13 scenario than in the 21-52-6-2-2-5 scenario because the evidence for their winning (21 tickets) was better than the evidence for the strongest alternative outcome in the first scenario (15 tickets) but not in the second scenario (52 tickets). In the present article, we refer to this comparison between the focal outcome and the strongest alternative as a *heuristic* comparison process because it is relatively efficient and serves as a roughly accurate guide to whether one should feel optimistic or pessimistic about the possibility of the focal outcome (see Windschitl & Wells, 1998).

This explanation for the alternative-outcomes effect does not assume that people are necessarily fooled or mistaken about the objective probability of a focal event. Rather, it assumes that there can be a dissociation between a person's belief in the objective probability of an event and his/her more intuitive or "gut-level" perceptions of certainty. That is, variations in how evidence is distributed across alternative outcomes can affect gut-level perceptions of certainty even if the variations have no influence on beliefs about the objective probability of the focal outcome. Recent research on other phenomena such as the ratio-bias phenomenon (Denes-Raj & Epstein, 1994; Kirkpatrick & Epstein, 1992) and context effects (Windschitl & Weber, 1999) has uncovered related evidence for dissociations between gut-level perceptions of certainty and beliefs in objective likelihood.

Consistent with this idea of a dissociation between gut-level perceptions of certainty and beliefs in objective probability, there is evidence of a dissociation in the types of measures that are sensitive to alternative-outcomes effects (Gonzalez & Frenck-Mestre, 1993; Teigen, 1988, in press; Windschitl & Wells, 1998). Verbal and other nonnumeric measures of likelihood have been used to detect alternative-outcomes effects, whereas numeric measures of subjective probability tend to show little sensitivity to the relevant manipulations. Windschitl and Wells (1996, 1998) argued that numeric measures of probability fail to detect such effects because the numeric measures—more so than nonnumeric ones—prompt people to apply their understandings of mathematical (normative) rules in search of the most accurate response. Because the present research was designed to investigate the influence of alternative outcomes on *gut-level* perceptions of certainty, we utilized a nonnumeric measure of certainty.

OVERVIEW OF THE EXPERIMENTS

Although the heuristic comparison hypothesis is consistent with known demonstrations of the alternative-outcomes effect, there are critical aspects of that hypothesis that warrant further testing. At the heart of the three experiments described here is the question of whether the primary cause of the alternative-outcomes effect is, as the hypothesis suggests, an influential comparison between the strength of evidence for the focal outcome and the strongest alternative. A key motivation for Experiment 1 was to test whether the rank-order status of the focal outcome (as the strongest or second strongest outcome) plays a primary role in the alternative-outcomes effect rather than the hypothesized comparison processes. Experiment 2 tested whether the hypothesized comparison processes would produce alternative-outcomes effects even when the focal outcome was weaker than all of the other alternatives. Experiment 3 tested the influence of the strongest alternative outcome relative to the other alternative outcomes. Not only do these experiments test aspects of the heuristic comparison account for alternative-outcomes effects, they also extend our understanding of the strength, the boundary conditions, and the generalizability of those effects.

The three experiments used the same basic methodology. Participants viewed numerous representations of raffles, which provided participants with information about how many tickets they held and how many tickets were held by each of the alternative players. For each raffle, participants indicated how good they felt about the possibility of winning. The dependent measure we used made it clear to the participants that we were interested in their gut-level estimates.

EXPERIMENT 1

The goals of Experiment 1 were threefold. First, we sought to develop a paradigm in which alternative-outcomes effects could be replicated and tested with several sets of distributions in a within-subjects design. This type of paradigm allows us to assess the generality and characteristics of the effect across numerous types of distributions, while holding constant the surface characteristics of the task, thus eliminating an additional source of variability present in earlier demonstrations of the effect. This paradigm also allows us to assess the consistency of the alternative-outcomes effect across individuals because each participant makes a large number of judgments. We can then determine whether the alternative-outcomes effects previously observed with between-subjects designs were produced by a minority of participants who were strongly affected by heuristic comparisons between the focal and strongest alternative outcomes or whether these comparison processes are characteristic of people in general.

A second goal of the experiment was to determine whether the magnitude of the alternative-outcomes effect is influenced by the *order* in which the alternative outcomes are described. In the scenario-based studies described by Windschitl and Wells (1998), the strongest alternative outcome was described immediately after the focal outcome was described. We assumed that this juxtaposition of the focal outcome and strongest alternative was not necessary for an alternative-outcomes effect. However, describing the strongest alternative immediately after the focal outcome may have selectively facilitated a contrastive comparison between the focal outcome and strongest alternative (as opposed to the focal and weaker alternatives that were not juxtaposed with the focal outcome) and consequently augmented the alternative-outcomes effects. Hence, it is important to manipulate description order to determine whether the alternative-outcomes effect partially depends on specific types of ordering or if the effect is robust to such a manipulation.

A third goal of the experiment was related to a key supposition of the heuristic comparison hypothesis. According to the hypothesis, the comparison between the focal and the strongest alternative outcome is based on their relative strengths, not simply a determination of which outcome is stronger (or more likely). This means that manipulating the distribution of alternative outcomes should affect certainty in the focal outcome even when the focal outcome's rank-order status is unaffected by the manipulation. Windschitl and Wells (1998) described one study in which this was true. In their Study 3, some participants read a 30-7-5-5-3 version of the critical scenario (with the "30" representing the strength of the focal outcome and the other numbers representing the strengths of alternatives) while others read a 30-20 version. Although the focal outcome was the strongest outcome in both versions, participants felt more optimistic about the focal outcome when reading the 30-7-5-5-3 version than the 30-20 version. However, in all other demonstrations of alternative-outcomes effects, the manipulations to the distributions of alternative outcomes also varied the rank order of the focal outcome (from strongest to second strongest). Hence, although Study 3 by Windschitl and Wells suggests that variations in rank-order status are not necessary for producing alternative-outcomes effects, additional demonstrations of this finding would certainly strengthen this critical conclusion. Furthermore, the methodology used in the present experiments allowed us to assess whether manipulations that do vary the rank-order status of the focal outcome are especially effective relative to those that do not. If so, then the heuristic comparison account should be modified to emphasize the importance of the focal outcome's rank-order status in producing alternative-outcomes effects.

Method

Participants. The participants were 96 students enrolled in introductory psychology courses at the University of Iowa.

Procedure. Participants were tested in groups ranging in size from 1 to 10. Each participant received a booklet that contained separate representations of 20 critical raffles as well as 6 filler raffles that are not discussed further. Figure 1 shows an example of the raffles seen by participants. A given subject saw these raffles in one of three random orders. To the right of each raffle, the following question appeared: "At a gut level, how would you feel about your chances of winning this raffle." The response scale was composed of nine equally spaced asterisks in a horizontal row with the anchor *not good at all* below the leftmost asterisk and the anchor *very good* below the rightmost asterisk. Participants circled an asterisk to indicate their response, and these responses were scored from 1 (*not good at all*) to 9 (*very good*). The initial instructions in the booklet explained how to use this scale and stressed to participants that: "We are interested in your initial impressions and your gut-level responses. We are *NOT* interested in your careful analysis of *exactly* how optimistic you *should* feel or in your precise assessments of the objective likelihood of winning."

In addition to manipulating the raffle distributions, we also manipulated the order in which the alternative players appeared in the raffle representations. For approximately half of the participants in the study, the strongest alternative player always appeared in the first position, to the immediate right of the focal player (labeled "you"; see Fig. 1), and the weakest alternative player always appeared in the fifth position, to the far right of the focal player. For the other half of the participants, the strongest alternative player always appeared in the third position, and the weakest appeared in the fourth position.

The raffles. Table 1 displays the raffles used in the experiment. As suggested by the table, the 20 critical raffles can be thought of conceptually as 5 sets of 4 raffles. In each set, there is a Baseline Raffle and three other raffles—Raffles A, B, and C.

The Baseline Raffles were constructed somewhat arbitrarily but with the constraint that the evidence for the focal outcome was far greater than the evidence for any single alternative outcome (i.e., the number of tickets held by the respondent was much greater than the number held by any alternative player). To construct Raffles A, B, and C in a given set, tickets were added to the Baseline Raffle of that set. To construct Raffle A, these added tickets were distributed across the four weakest alternative players (with one small exception in Set 2). To construct Raffle B, the same number of tickets were added primarily to the strongest alternative player. To construct Raffle C, the same number of tickets were added to the strongest alternative player and some tickets were redistributed from the weaker players and concentrated in the strongest alternative player. Consider, for example, Set 2. The total number of tickets in the Baseline Raffle of Set 2 was 48, whereas the number of tickets in Raffles A, B, and C was 59 each. Hence, the objective probability of winning was higher in the Baseline Raffle than in Raffles A, B, and C, but the objective probability of winning did not differ among Raffles A, B, and C. The only difference between Raffles A, B, and C concerned the manner in which evidence

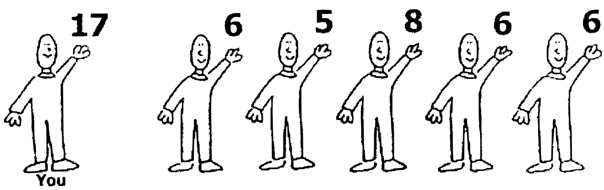

FIG. 1. An example of the types of raffles seen by participants in Experiments 1, 2, and 3.

TABLE 1

Information about the Raffles Used in Experiment 1

Raffle name	Tickets held by players						Total	*P*(Focal)
Set 1								
Baseline	14	5	3	2	2	2	28	0.50
A	14	5	5	5	5	5	39	0.36
B	14	13	3	3	3	3	39	0.36
C	14	20	2	1	1	1	39	0.36
Set 2								
Baseline	17	8	6	6	6	5	48	0.35
A	17	9	9	8	8	8	59	0.29
B	17	16	7	7	6	6	59	0.29
C	17	32	3	3	2	2	59	0.29
Set 3								
Baseline	18	6	4	3	3	3	37	0.49
A	18	6	6	6	6	6	48	0.38
B	18	17	4	3	3	3	48	0.38
C	18	22	3	2	2	1	48	0.38
Set 4								
Baseline	21	9	7	6	6	6	55	0.38
A	21	9	9	9	9	9	66	0.32
B	21	20	7	6	6	6	66	0.32
C	21	25	6	5	5	4	66	0.32
Set 5								
Baseline	32	19	12	12	11	10	96	0.33
A	32	19	18	18	17	16	120	0.27
B	32	30	15	15	14	14	120	0.27
C	32	43	12	12	11	10	120	0.27

Note. The underlined column indicates the numbers of tickets held by the respondent in each raffle (the focal outcome). The remaining ticket columns indicate the numbers of tickets held by other individual players (ordered according to ticket amount). The two columns on the far right indicate the total number of tickets specified in each raffle and the objective probability that the respondent would win the raffle (i.e., the probability of the focal outcome).

was distributed across the alternative outcomes. These properties held true for the raffles within each of the five sets.

Results and Discussion

If people are sensitive only to comparisons between the evidence for the focal outcome and the aggregated evidence for the alternative outcomes, then the only differences in certainty that we should expect within a raffle set would be between the Baseline Raffle and the A, B, and C Raffles. However, we expected that participants would also be sensitive to how the evidence was distributed across the alternative outcomes (consistent with the effects demonstrated by Windschitl & Wells, 1998). Specifically, we expected that comparisons between the focal and the strongest alternative outcome would have a particularly strong influence on participants' gut-level perceptions of certainty. Hence, for Raffles A, B, and C, we predicted that certainty in the focal outcome would be greatest in the A Raffles (because the evidence for the focal outcome was much stronger than the evidence for the strongest alternative) and weakest in the C Raffle (because the strongest alternative was actually stronger than the focal outcome). This result would be a conceptual replication of the effects detected in the three scenario-based studies described by Windschitl and Wells (1998) that manipulated both the distribution of alternative outcomes and the rank-order status of the focal outcome. The results for the B Raffles offer an opportunity to assess whether the rank-order status of the focal outcome is an especially important determinant of the outcome's perceived certainty and of the magnitude of an alternative-outcomes effect. If participants' certainty for the B Raffles were the same as that for the A Raffles but significantly more than that for the C Raffles, this would suggest that variations in the rank-order status of the focal outcome play a special role in the alternative-outcomes effects. However, consistent with the findings of Study 3 by Windschitl and Wells (1998), we expected to observe a significant difference in certainty between Raffles A and B. We also suspected that shifting tickets to the strongest alternative outcome would have roughly the same impact on perceived certainty regardless of whether the shift produced a change in the rank-order status of the focal outcome. Hence, we expected that the overall difference in perceived certainty between Raffles A and B would be roughly the same as the overall difference for Raffles B and C.

The results were quite consistent with these predictions as well as our expectation that the order in which the alternative outcomes were shown would have little impact on alternative-outcomes effects. Table 2 displays the means of the certainty responses for each raffle of each set as well as means based on the responses across all Baseline Raffles, across all A Raffles, and so on. Participants' responses were initially analyzed in a repeated-measures MANOVA with raffle type (Baseline, A, B, or C) and raffle set (1, 2, 3, 4, or 5) as within-subjects factors and display order (juxtaposition/nonjuxtaposition of focal and strongest alternatives) as a between-subjects factor. Neither the main effect nor any interactions involving the display-order factor were significant. Hence, this factor was dropped for all further analyses.

Not surprisingly, the main effect for raffle set was significant, $F(4, 91) = 14.86$, $p < .001$. The Raffle Set × Raffle Type interaction was also significant, $F(12, 83) = 6.05$, $p < .001$. Most importantly, the main effect for raffle type was significant, $F(3, 92) = 75.32$, $p < .001$. Planned univariate contrasts on the mean responses to raffle types (collapsed across raffle sets) revealed that participants' perceptions of certainty were significantly different between all possible pairs of raffle types (all *p*s $< .001$). Analogous pairwise comparisons within each raffle set suggested that this overall pattern was generally characteristic of all the sets (see Table 2 for information about these comparisons).

The results of these analyses reveal several important discoveries. First, the alternative-outcomes effect can be reliably replicated with this raffle paradigm across a variety of numeric distributions. There is no evidence to suggest that the effect is restricted to an idiosyncratic set of distributions; the comparisons

TABLE 2

Mean Certainty Responses for Each Raffle in Experiment 1

Raffle name	*P*(Focal)	Mean	*SD*
Set 1			
Baseline	0.50	7.47	1.46
A	0.36	6.78[a]	1.80
B	0.36	6.60[a]	1.59
C	0.36	5.60	1.65
Set 2			
Baseline	0.35	6.97[a]	1.70
A	0.29	6.67[a]	1.67
B	0.29	5.79	1.76
C	0.29	4.67	1.80
Set 3			
Baseline	0.49	7.37	1.60
A	0.38	6.82	1.86
B	0.38	6.19	1.84
C	0.38	5.74	1.63
Set 4			
Baseline	0.38	6.86[a]	1.72
A	0.32	6.94[a]	1.81
B	0.32	5.76[b]	1.71
C	0.32	5.88[b]	1.80
Set 5			
Baseline	0.33	6.59[a]	2.09
A	0.27	6.39[a]	2.16
B	0.27	5.69	1.92
C	0.27	4.57	1.84
Overall			
Baseline	0.41	7.03	1.42
A	0.32	6.72	1.50
B	0.32	6.00	1.48
C	0.32	5.29	1.38

Note. The column labeled "*P*(Focal)" indicates the objective probability that the respondent would win the raffle (i.e., the probability of the focal outcome). Certainty responses were scored on a scale from 1 to 9. Within each raffle set, means that share superscripts were not significantly different from each other; all other pairwise comparisons within a given set were significant ($p < .05$).

between the A and C Raffles were significant for all five of the sets. Hence, it seems reasonable to argue that the comparison processes producing the effect are not triggered on a case-by-case basis (i.e., triggered by some aspect of a particular distribution of evidence), but rather constitute basic processes by which people's gut-level perceptions of certainty are determined.

Second, it is also important to note that participants were sensitive to the differences in the objective probabilities of focal outcomes (e.g., between the Baseline Raffles and the A Raffles), even when those differences in perceived certainty could not be attributed to comparisons between the focal and strongest alternative outcomes. In other words, adding tickets to nonfocal outcomes had a significant effect on the perceived certainty of the focal outcome, even when those tickets were added to players that were weaker than the strongest alternative player. Hence, although the heuristic comparison hypothesis suggests that the comparison between the focal and strongest alternative outcome is the primary determinant of alternative-outcomes effects, it would be inaccurate to argue that participants only attend to the focal and strongest alternative; they are at least somewhat sensitive to evidence supporting the weaker alternatives.

Third, the fact that the magnitude of alternative-outcomes effects was not significantly influenced by the positioning of the focal and strongest alternative outcomes attests to the robustness of the processes underlying alternative-outcomes effects. It appears that regardless of where the strongest alternative outcome is displayed relative to the focal outcome, it has a greater influence on the perceived certainty of the focal outcome than do weaker alternative outcomes.

Fourth, the results provide compelling evidence that variations in the rank-order status of the focal outcome are not necessary for producing an alternative-outcomes effect. In the A Raffles, the focal outcome was much stronger than all alternatives, whereas in the B Raffles, the strongest alternative outcome approached, but did not exceed, the strength of the focal outcome. Even though the focal outcome was the strongest outcome in both the A and B Raffles of every set, an alternative-outcomes effect (i.e., a significant difference between Raffles A and B) was observed in four of the five sets. Furthermore, it is instructive to note that the overall difference in perceived certainty between the A and B Raffles was approximately the same as the overall difference between the B and C Raffles ($t = 0.0$ for a test comparing the two difference scores), even though the rank-order status of the focal outcomes varied between the B and C Raffles. The overall differences between Raffles A and B and between B and C appear to be attributable to the fact that the strongest alternative player had, on average, 9.6 more tickets in Raffle B than in A and 9.4 more tickets in Raffle C than in B. There does not appear to be a notable magnification of the alternative-outcomes effect when the rank-order status of the focal outcome shifts. Hence, not only is a manipulation of rank-order status not necessary for observing alternative-outcomes effects, rank-order variations per se do not appear to play a particularly strong role in producing such effects.

Fifth, the data provide evidence relevant to questions about the generalizability of the effect across individuals. To assess this generalizability, we determined the numbers of participants whose mean response to the A Raffles was greater than, equal to, or less than their mean response to the C Raffles. Note that if a given participant used the type of comparison processes described above, then his/her responses would be greater for the A Raffles than for the C Raffles. Was this pattern of responding generally consistent across individuals, or was there a notable portion of participants who showed the reverse pattern? For 84% of the participants, the mean of their responses to the A Raffles was higher than the mean of their responses to the C Raffles. For the remaining 5 and

9% of the participants, the mean of their responses was equal or less, respectively, for the A Raffles than for the C Raffles.[2] These results clearly indicate that the alternative-outcomes effect cannot be attributed to the response patterns of only a minority of participants. The comparison processes mediating the alternative-outcomes effects appear to be influential not only across various distributions but across individuals as well.

Finally, Experiment 1 provides information relevant to the question of how powerful the comparison processes that produce the alternative-outcomes effects can be. Cohen's (1988) effect size estimate (standardized mean difference) was 0.99 for the difference between the A and C Raffles, and it was 0.48 for the difference between the A and B Raffles. Note that the difference in the objective probability of winning among these raffles was 0%. By comparison, the average difference in the objective probability of winning between the Baseline and A Raffles was 9%, whereas the effect size estimate for the difference between these raffles was only 0.21. Hence, manipulations to the evidence distributions had a substantially stronger impact on perceptions of certainty than did the manipulations of objective probability for the distributions used in Experiment 1.

EXPERIMENT 2

In each of the distributions tested above and in those tested in other research, the focal outcome was one of the strongest of all possible outcomes. This reflects some ecological validity in the sense that, in everyday life, likely outcomes are thought about more often than unlikely ones. However, there are certainly times at which we are keenly interested in the possibility of a relatively unlikely outcome. When thinking about a focal outcome that is weak, is a comparison to the strongest alternative outcome still particularly influential? It is possible that when the focal outcome is weak, comparisons to weaker alternative outcomes become more important. If evidence in weaker alternatives has the same influence or more influence than evidence in the stronger alternatives, then alternative-outcomes effects would be neutralized or reversed in direction. For example, it is possible that a raffle player who holds only 3 tickets in a raffle might feel more certain about winning if his/her three opponents hold 10, 4, and 4 tickets than if his/her three opponents hold 6, 6, and 6 tickets; in the former case, but not the latter, the focal outcome is almost as strong as at least one other outcome. This possibility that alternative-outcomes effects would neutralize or reverse in direction when the focal outcome is weak was investigated in Experiment 2.

Despite the plausible rationale for a reversal of the alternative-outcomes effect, we expected that, consistent with the heuristic comparison hypothesis, the strongest alternative outcome would play a primary role in determining

[2] The failure of these three numbers to add to 100 is due to rounding error. Analogous values that reflect the percentage of participants whose mean responses to the A Raffles were higher than, equal to, or lower than their mean responses to the B Raffles were 73, 7, and 20%.

perceptions of certainty regardless of whether the focal outcome was weak or strong. The raffle player holding 3 tickets should feel more certain about winning when the opponents hold 6, 6, and 6 tickets than when they hold 10, 4, and 4 tickets because the comparison between the focal outcome and strongest alternative is less *un*favorable in the 3-6-6-6 situation than in the 3-10-4-4 situation. If, however, a player who holds relatively few tickets becomes particularly sensitive to how likely/unlikely his or her winning is relative to players in a similar situation (i.e., those holding few tickets), the direction of the alternative-outcomes effect would be reversed, and such a reversal would lead to an important qualification of the hypothesized role of the strongest alternative outcome in producing alternative-outcomes effects.

Method

Participants. A separate sample of 72 students, enrolled in introductory psychology courses at the University of Iowa, served as participants for Experiment 2.

Procedure. The procedures were identical to those of Experiment 1 with one exception. We again manipulated the order in which the alternative players were displayed in the raffles; however, this between-subjects manipulation was slightly more elaborate in Experiment 2. Each participant saw some raffles in which the strongest alternative player appeared in the second position and some raffles in which the strongest alternative player appeared in the fourth position. Whether a participant saw a “second-position” or “fourth-position” version for a given raffle was determined by the between-subjects manipulation. As was the case for Experiment 1, the position factor did not interact with the results described below and hence is not discussed further.

The raffles. Table 3 displays the raffles used in the experiment. As suggested by the table, the 18 critical raffles can be thought of conceptually as 6 sets of 3 raffles. As in Experiment 1, the Baseline Raffles were constructed somewhat arbitrarily but with the new constraint that the evidence for the focal outcome should be weaker than at least two of the alternative outcomes. In fact, the focal outcome was the weakest outcome in 4 of the 6 Baseline Raffles. To construct Raffles A and B in a given set, tickets were added to the Baseline Raffle from that set. For Raffle A, the added tickets were distributed across the four weakest alternative players. For Raffle B, the same number of tickets were instead added to the strongest alternative player. Hence, similar to Experiment 1, the objective probability of winning was higher in the Baseline Raffle than in Raffles A and B, but the probability of winning did not differ between Raffles A and B.

Results and Discussion

Table 4 displays mean certainty responses. Participants' responses were analyzed in a repeated-measures MANOVA with raffle type and raffle set as

TABLE 3

Information about the Raffles Used in Experiment 2

Raffle name	Tickets held by players						Total	*P*(Focal)
Set 1								
Baseline	10	15	12	12	12	11	72	0.14
A	10	15	14	14	14	13	80	0.13
B	10	23	12	12	12	11	80	0.13
Set 2								
Baseline	6	10	8	7	7	7	45	0.13
A	6	10	9	9	9	8	51	0.12
B	6	16	8	7	7	7	51	0.12
Set 3								
Baseline	11	12	12	7	6	6	54	0.20
A	11	12	12	9	9	8	61	0.18
B	11	19	12	7	6	6	61	0.18
Set 4								
Baseline	21	30	24	23	23	22	143	0.15
A	21	30	29	28	28	27	163	0.13
B	21	50	24	23	23	22	163	0.13
Set 5								
Baseline	9	11	11	11	3	3	48	0.19
A	9	11	11	11	7	7	56	0.16
B	9	19	11	11	3	3	56	0.16
Set 6								
Baseline	19	39	23	23	22	22	148	0.13
A	19	39	38	38	37	36	207	0.09
B	19	98	23	23	22	22	207	0.09

Note. The underlined column indicates the numbers of tickets held by the respondent in each raffle (the focal outcome). The remaining ticket columns indicate the numbers of tickets held by other individual players (ordered according to ticket amount). The two columns on the far right indicate the total number of tickets specified in each raffle and the objective probability that the respondent would win the raffle (i.e., the probability of the focal outcome).

factors. The main effect for raffle set was again significant, $F(5, 67) = 76.58$, $p < .001$, as was the Raffle Set x Raffle Type interaction, $F(10, 62) = 2.15$, $p < .05$. More importantly, the main effect for raffle type was significant, $F(5, 67) = 57.15$, $p < .001$. Planned univariate contrasts on the mean responses to raffle types (collapsed across raffle sets) revealed that participants' perceptions of certainty were significantly different between all possible pairs of raffle types (all *p*s $< .01$), including the critical comparison for detecting an alternative-outcomes effect between Raffles A and B.

Hence, even when the focal outcome was relatively weak, evidence that was added to the strongest alternative had greater impact on the perceived certainty of the focal outcome than did evidence that was added across weaker alternatives. It is clear from Table 4, however, that the alternative-outcomes effects observed in Experiment 2 were weaker than those in Experiment 1. The effect size estimate for the difference between Raffles A and B in Experiment 2 was only 0.25 compared to the 0.99 and 0.48 estimates mentioned above for

TABLE 4

Mean Certainty Responses for Each Raffle in Experiment 2

Raffle name	*P*(Focal)	Mean	*SD*
Set 1			
Baseline	0.14	3.58	1.76
A	0.13	3.08[a]	1.57
B	0.13	2.82[a]	1.17
Set 2			
Baseline	0.13	3.49	1.59
A	0.12	2.67[a]	1.35
B	0.12	2.86[a]	1.58
Set 3			
Baseline	0.20	4.96[a]	1.60
A	0.18	5.00[a]	1.61
B	0.18	4.50	1.48
Set 4			
Baseline	0.15	3.40	1.82
A	0.13	2.90	1.39
B	0.13	2.44	1.24
Set 5			
Baseline	0.19	4.54	1.69
A	0.16	4.02[a]	1.49
B	0.16	3.69[a]	1.40
Set 6			
Baseline	0.13	2.44	1.25
A	0.09	1.94[a]	1.12
B	0.09	1.68[a]	1.20
Overall			
Baseline	0.16	3.99	1.35
A	0.13	3.54	1.19
B	0.13	3.26	1.02

Note. The column labeled "*P*(Focal)" indicates the objective probability that the respondent would win the raffle. Within each raffle set, means that share superscripts were not significantly different from each other; all other pairwise comparisons within a given set were significant ($p < .05$). The p values for the *non*significant comparisons in Sets 1, 5, and 6, were $<.10$.

Experiment 1. One reason why the alternative-outcomes effects were smaller in Experiment 2 may have been that the manipulations of the strengths of the strongest alternative outcomes were not as severe in Experiment 2 as they were in Experiment 1. However, the reduced size of the alternative-outcomes effects may also signal that the use of the comparison heuristic (or in other words, the influence of the comparison between the focal outcome and strongest alternative) may be moderated by the general strength of the focal outcome. This issue is addressed again under General Discussion, but the key point here is that manipulations of evidence across alternative outcomes can influence the perceived certainty of the focal outcome in a direction consistent with the heuristic comparison account, even when the focal outcome is quite weak.

EXPERIMENT 3

In Experiments 1 and 2, tickets that were added to the alternative outcomes had a different impact on perceived certainty in the focal outcome depending on how those added tickets were distributed. The tickets were significantly more influential when they were added to the strongest alternative than when they were distributed across weaker alternatives. Throughout this article, we have attributed these effects to a heuristic comparison process which assumes that the strongest alternative outcome plays a much greater role in producing alternative-outcome effects than does any other single alternative. However, although Experiments 1 and 2 as well as other demonstrations of alternative-outcomes effects (e.g., Windschitl & Wells, 1998) are fully consistent with this heuristic comparison account, none of these experiments have directly tested the claim that that the strongest alternative outcome plays a dominant role.

There are at least two ways of explaining the alternative-outcomes effects described here and by Windschitl and Wells (1998) that do not assume any special status for the strongest alternative per se. One explanation is that there is a limit (or soft constraint) to the number of outcomes that people will consider when making a gut-level estimate. Perhaps, for example, people consider evidence only for the three strongest alternative outcomes, not all five that were presented in the raffles of Experiments 1 and 2. Hence, any evidence that is added to the two weakest alternatives is “wasted” in the sense that it will not impact certainty in the focal outcome. Another explanation is that when added evidence is distributed across many outcomes, it is less influential than when that evidence is concentrated on only one outcome, even when this outcome was not the strongest one. We conducted Experiment 3 to evaluate these alternative hypotheses and determine whether the strongest alternative outcome per se is significantly more influential than any other alternative outcome for producing alternative-outcomes effects.

Determining whether the strongest alternative outcome plays a dominant role in producing alternative-outcomes effects requires the detection of robust alternative-outcomes effects. In the raffles of Experiment 3, the focal outcome was always the most likely outcome; the results of Experiments 1 and 2 suggest that alternative-outcomes effects might be most prominent in such situations. Like Experiments 1 and 2, Experiment 3 manipulated the way in which raffle tickets were distributed among alternative outcomes. In Experiments 1 and 2, these added tickets were either distributed broadly or were generally concentrated into the strongest alternative. Experiment 3 followed the same pattern, but also included distributions in which the added tickets were concentrated into the *second strongest* alternative outcome. Hence, the critical question was whether adding tickets to the second strongest alternative outcome had the same effect on certainty in the focal outcome (i.e., produced the same magnitude of an alternative-outcomes effect) as did adding those tickets to the strongest alternative outcome. Consistent with our formulation of the heuristic comparison hypothesis, and in opposition to the two alternative hypotheses mentioned above, we expected that adding tickets to the strongest alternative outcome would have substantially more impact than adding the same number of tickets to the second strongest alternative outcome.

Method

Participants. A separate sample of 72 students, enrolled in introductory psychology courses at the University of Iowa, served as participants

Procedure. The procedures were identical to those of Experiment 2. We again manipulated the order in which the alternative players appeared in the raffle dislays, but this position factor did not interact with the results described below and hence will not be discussed further.

The raffles. Table 5 displays the 16 raffles (4 sets of 4) used in the experiment. As in Experiment 1, the Baseline Raffles were constructed somewhat arbitrarily but with the constraint that the focal outcome was stronger than all alternatives. To construct the A Raffles, additional tickets were distributed across the four weakest alternative players. To construct the B Raffles, the same number of tickets were instead added to the *second-strongest* alternative

TABLE 5

Information about the Raffles Used in Experiment 3

Raffle name	Tickets held by players						Total	*P*(Focal)
Set 1								
Baseline	19	11	3	2	2	1	38	0.50
A	19	11	4	4	4	2	44	0.43
B	19	11	9	2	2	1	44	0.43
C	19	17	3	2	2	1	44	0.43
Set 2								
Baseline	30	17	5	5	4	3	64	0.47
A	30	17	8	7	7	6	75	0.40
B	30	17	16	5	4	3	75	0.40
C	30	28	5	5	4	3	75	0.40
Set 3								
Baseline	25	16	8	8	8	7	72	0.35
A	25	16	10	10	10	9	80	0.31
B	25	16	16	8	8	7	80	0.31
C	25	24	8	8	8	7	80	0.31
Set 4								
Baseline	14	8	3	3	2	2	32	0.44
A	14	8	4	4	4	3	37	0.38
B	14	8	8	3	2	2	37	0.38
C	14	13	3	3	2	2	37	0.38

Note. The underlined column indicates the numbers of tickets held by the respondent in each raffle (the focal outcome). The remaining ticket columns indicate the numbers of tickets held by other individual players (ordered according to ticket amount). The two columns on the far right indicate the total number of tickets specified in each raffle and the objective probability that the respondent would win the raffle (i.e., the probability of the focal outcome).

player. Note that the addition of these tickets to the second-strongest alternative player did not change the status of this player; this player remained the second-strongest alternative. To construct the C Raffles, the same number of tickets were instead added to the *strongest* alternative player. Note again, that this addition did not change the status of that player (from weaker to stronger than the focal player). Hence, as in Experiment 1, the objective probability of winning was higher in the Baseline Raffle than in Raffles A, B, and C, but the objective probability of winning did not differ among Raffles A, B, and C.

Results and Discussion

Table 6 shows the mean certainty responses. Participants' responses were analyzed using a repeated-measures MANOVA with raffle type and raffle set as factors. Not surprisingly, the main effect for raffle set was again significant, $F(3, 69) = 42.04$, $p < .001$. The Raffle Set × Raffle Type interaction was not significant, $F < 1$. Most importantly, the main effect for raffle type was

TABLE 6

Mean Certainty Responses for Each Raffle in Experiment 3

Raffle name	*P*(Focal)	Mean	SD
Set 1			
Baseline	0.50	7.17[ab]	1.32
A	0.43	7.26[a]	1.33
B	0.43	6.89[b]	1.41
C	0.43	6.25	1.41
Set 2			
Baseline	0.47	7.63[a]	1.28
A	0.40	7.51[b]	1.35
B	0.40	6.85	1.74
C	0.40	6.40	1.62
Set 3			
Baseline	0.35	6.63[a]	1.38
A	0.31	6.36[a]	1.81
B	0.31	6.49[a]	1.46
C	0.31	5.44	1.68
Set 4			
Baseline	0.44	7.13[a]	1.44
A	0.38	6.97[ab]	1.50
B	0.38	6.69[b]	1.56
C	0.38	6.18	1.59
Overall			
Baseline	0.44	7.14	1.02
A	0.38	7.03	1.08
B	0.38	6.73	1.11
C	0.38	6.07	1.24

Note. The column labeled "*P*(Focal)" indicates the objective probability that the respondent would win the raffle (i.e., the probability of the focal outcome). Certainty responses were scored on a scale from 1 to 9. Within each raffle set, means that share superscripts were not significantly different from each other; all other pairwise comparisons within a given set were significant ($p < .05$).

significant, $F(3, 69) = 32.22$, $p < .001$. Planned univariate contrasts on the mean responses to raffle types (collapsed across raffle sets) revealed that participants' perceptions of certainty were significantly different between all possible pairs of raffle types (all ps $< .001$) with one exception: the mean responses to the Baseline Raffles and the A Raffles were not significantly different ($p = .24$). Hence, unlike Experiments 1 and 2, participants did not show significant sensitivity to the decrease in objective probability that resulted from distributing additional tickets among the weaker alternative players.

The significant difference between the A and the B Raffles is instructive in that it demonstrates that evidence added to the second-strongest alternative outcome has more impact than evidence distributed across weaker alternative outcomes. The most important comparison for this experiment was between the B Raffles and the C Raffles. As predicted, participants expressed more certainty for the B Raffles (in which the additional tickets were added to the second-strongest alternative) than for the C Raffles (in which the additional tickets were added to the strongest alternative). This finding rules out the two alternative hypotheses described above. First, if participants had attended equally to the three strongest alternatives and ignored the rest, we would not have observed significant differences between the B and C Raffles. Second, if it is the concentration (vs broader distribution) of added tickets that produces the alternative-outcomes effects in this paradigm, we would have observed no differences between the B and C Raffles. The observed pattern of results supports the claim that the strongest alternative outcome is indeed the most influential alternative outcome.

How strong was the impact of the strongest alternative relative to the impact of the second-strongest alternative? Adding tickets to the second-strongest alternative decreased the mean certainty response by 0.41 points (i.e., the difference between the mean responses for the Baseline Raffles and the B Raffles). Adding the same number of tickets to the strongest alternative decreased the mean certainty response by 1.07 points (i.e., the difference between the mean responses for the Baseline Raffles and the C Raffles). Hence, in this experiment the added tickets had over twice the impact (2.6 times the impact) when those tickets were located in the strongest rather than the second-strongest alternative outcome.

GENERAL DISCUSSION

In addition to establishing a paradigm for reliably replicating the alternative-outcomes effect and illustrating its generalizability across people and evidence distributions, the present experiments ruled out possible alternatives to the heuristic comparison account and provided some indication of the role that related processes play (or do not play) in producing such effects. Experiment 1 demonstrated that manipulations of the rank-order status of the focal outcome (e.g., from first to second strongest) are not necessary for observing alternative-outcomes effects, nor do such manipulations provide a special boost in the magnitude of the effects (beyond what can be accounted for by the heuristic comparisons). The results also illustrate that the heuristic comparisons are largely impervious to superficial manipulations of whether the focal and strongest alternative outcomes were presented in juxtaposed positions. Experiment 2 demonstrated that, even when the focal outcome is quite weak, evidence supporting the strongest alternative outcome is more influential than evidence distributed across weaker alternatives. Experiment 3 ruled out any accounts that did not assume that the strongest alternative outcome has a stronger influence in producing alternative-outcomes effects than do other alternative outcomes. In that experiment, evidence supporting the strongest alternative had a much greater impact on the perceived certainty of the focal outcome than did evidence supporting the second strongest alternative.

The results of the experiments are consistent with the hypothesis that a comparison between the focal and strongest alternative outcome plays a dominant role in producing alternative-outcomes effects. What, more specifically, constitutes a comparison between the focal and strongest alternative? We conceptualize this comparison as an assessment of relative strengths. Assume that F represents the strength of the evidence for the focal outcome and that A_j represents the strength of the evidence for the j^{th} strongest alternative. If only a heuristic comparison between the focal outcome and strongest alternative determined the perceived certainty of the focal outcome, then the perceived certainty of the focal outcome could be represented as $(F - A_1)/(F + A_1)$. Note that this model, which we will call the comparison-to-the-strongest model, produces predictions that are perfectly correlated with the simpler model: $F/(F + A_1)$. Although we henceforth use the simpler model as our representation of the comparison-to-the-strongest model, we think it is instructive to start with a description of the first equation because it more intuitively represents that the key process that determines gut-level perceptions of certainty is a comparison between the focal and strongest alternative outcome. Both of these representations are importantly distinct from a difference model: $F - A_1$. In other words, it is not the absolute difference between the amounts of evidence supporting the focal and strongest alternative that shapes perceived certainty, but the difference between these amounts relative to the overall amount of evidence relevant to these two outcomes. This means that, for example, in a raffle in which the focal and strongest alternative player each hold 10 tickets, adding 10 more tickets to the strongest alternative player would have a large impact, but in a raffle in which the two players each hold 1000 tickets, adding 10 more tickets to the strongest alternative player would have a minimal impact.

Whereas $F/(F + A_1)$ would represent a heuristic approach to determining perceived certainty in the focal outcome, $F/(F + A_1 + A_2 + A_3 + A_4 + A_5)$ would represent a purely normative approach to determining perceived certainty in the focal outcome. In the purely normative model, evidence for all alternative outcomes is fully considered. By conducting regression analyses of the mean participant ratings for each raffle within an experiment, we can evaluate the predictive efficacy of the comparison-to-the-strongest model relative to the normative model. This type of analysis can provide a rough assessment of the relative impact of heuristic versus normative processes within an experiment and how the relative impact of these processes differed across experiments.[3]

For the data from Experiment 1, when the normative (objective-probability) model was the sole predictor variable, it accounted for only 39% of the variance in mean ratings. In contrast, when we used the comparison-to-strongest model as the sole predictor variable, it accounted for 82% of the variance in the mean ratings. The analogous values for Experiment 2 were 91% for the normative model and 60% for the comparison-to-the-strongest model and for Experiment 3 were 40 and 63%.

In a regression analysis that included the normative model as well as the comparison-to-the-strongest model as predictor variables for the data in Experiment 1, each uniquely accounted for 5 and 45% of the variance, respectively, with 38% shared variance.[4] The respective values for Experiment 2 were 32 and 2%, with 60% shared variance, and for Experiment 3 were 19 and 40%, with 25% shared variance.

These analyses are consistent with the conclusions drawn from the MANOVA results described above—the impact of the comparison between the focal and strongest alternative outcome can be quite strong relative to the impact of other processes. However, the analyses also suggest that the influence of the heuristic comparison fluctuated across experiments. For Experiments 1 and 3, the heuristic comparison between the focal and strongest alternative outcome appears to have had a dominating influence on the perceived certainty of the focal outcome. For Experiment 2, although the MANOVA results clearly revealed significant alternative-outcomes effects, the regression analyses suggest that the role of the hypothesized comparison heuristic did not dominate over processes described by the normative model.

Determining why the comparison heuristic played a greater role in Experiments 1 and 3 than in Experiment 2 is worthy of future research. We mentioned earlier that one potential explanation for the differential effects between Experiments 1 and 2 was that the manipulations of the strengths of the strongest alternative outcomes were not as severe in Experiment 2 as in Experiment 1. However, the fact that analogous manipulations in Experiment 3 were no stronger than they were in Experiment 2, yet produced more robust alternative-outcomes effects, does not bode well for that explanation. Another possibility,

[3] We assume that the best fitting model for data from a given experiment would be one in which each alternative outcome would be individually weighted, with the strongest receiving the most weight and the weakest receiving the least weight (but not zero weight). However, from the limited data of our experiments, it would be impossible to know whether this type of model would accurately reflect the processes used by individual participants or whether this type of model produces a good fit because some participants primarily used a heuristic strategy, whereas others primarily used a normative strategy. While this issue should be addressed in future research, the main focus of the regression analyses we present is to contrast the extent to which a purely normative approach can account for the data, versus the extent to which the hypothesized comparison heuristic can account for the data.

[4] The unique variance accounted for is equivalent to the semipartial correlation squared (Cohen & Cohen, 1983).

mentioned earlier, is that there is an increase in a respondent's interest in weaker alternative outcomes when the focal outcome is also weak (although this increase is not expected to eclipse the role of the strongest alternative outcome). Another related but distinct possibility is that being in a weak position motivates respondents to more carefully evaluate their chances and thus makes them more likely to employ an aggregation strategy. Other possible explanations exist as well, and testing them may have some important implications for understanding the moderators of alternative-outcomes effects inside as well as outside this raffle paradigm.

Comparison-to-the-Strongest as a Heuristic

We argue that comparing the focal outcome to the strongest alternative can serve as a useful heuristic—that is, a process that usually yields good results with minimal effort. Like other recently proposed heuristics that suggest that people sometimes use one or two simple but generally valid cues to make seemingly complex decisions (e.g., take-the-best; Gigerenzer & Goldstein, 1999), the comparison heuristic suggests that people's perceptions of certainty are often heavily or perhaps sometimes exclusively determined by consideration of only two pieces of information. A participant's use of such a heuristic in these experiments could obviate the need to aggregate the evidence that was distributed across the six possible outcomes. Instead, the participant could simply find the strongest alternative and compare its support relative to the support for the focal outcome. In these experiments, we manipulated the evidence distributions so that the use of the heuristic would produce detectable, nonnormative response patterns. In most natural contexts, however, we suspect that such a heuristic would likely provide a roughly accurate estimate of probability. This supposition, however, requires a different form of empirical test (for related discussions see Gigerenzer & Todd, 1999, Gigerenzer & Goldstein, 1999).

Is the Comparison Heuristic Relevant Only to Gut-Level Perceptions?

The present research has focused on gut-level perceptions of certainty. However, we do not wish to suggest that the comparison heuristic is relevant only for gut-level perceptions that are solicited with scales like those used here. There is clear evidence that gut-level perceptions measured with nonnumeric scales mediate other judgments, decisions, and behavior. Windschitl and Wells (1996) demonstrated that responses on verbal scales (the same scales that they later used to investigate alternative-outcomes effects; Windschitl & Wells, 1998) can be better predictors of people's preferences and behavioral intentions under uncertainty than are responses on numeric subjective probability scales. That is, relative to their numeric responses, participants' verbal responses about the likelihood of a critical outcome for a described situation were better predictors of what the participants reported they would do if they were in that situation. Also, in studies that used judgments (Study 3; Windschitl & Wells, 1998) and choices (Study 5) as the primary dependent measures rather than verbal or numeric likelihood estimates, variations in the distributions of alternative outcomes nevertheless had robust effects. For example, in a study in which participants were offered the choice of a ticket in two formally equivalent raffles, participants tended to select the raffle in which no other individual player held more than one ticket rather than the raffle in which one other player held several tickets.

While the heuristic comparison processes described here are likely to be influential in a variety of contexts, we assume that the heuristic processes are usually partial—not complete—determinants of a given likelihood judgment. The relative influence of these comparison processes and more deliberative aggregation processes likely depends on a number of task characteristics. For example, we suspect that as the evidence for the possible outcomes becomes more difficult to aggregate, reliance on heuristic comparisons will increase. Also, although we stated earlier that numeric judgments of probability are less likely to exhibit evidence of heuristic comparisons than are nonnumeric judgments, there may be situations in which the comparison heuristic is influential even if respondents are trying to assess objective likelihood on a numeric scale. For example, if the evidence is especially difficult to aggregate and/or a respondent has only a limited amount of time in which to provide his/her best estimate of an event's likelihood, he/she may be wise to consider only the evidence for the focal event and the strongest alternative event rather than attempt to assess the evidence for all possible alternatives.

Implications for Broader Theory and Related Phenomena

The present findings suggest some novel considerations for theories of likelihood judgment. Most relevant is support theory, a recent and prominent theory of subjective probability (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994). A central assumption of support theory is that the judged probability, $P(A, \overline{A})$, that a focal hypothesis holds rather than any of its alternatives, which are collectively referred to as the residual hypothesis, depends on support for the focal hypothesis, $s(A)$, and the residual hypothesis, $s(\overline{A})$, as follows:

$$P(A, \overline{A}) = \frac{s(A)}{s(A) + s(\overline{A})} .$$

Building from this framework, support theory has accounted for a variety of phenomena, including people's tendency to overestimate the probability of focal hypotheses such that their estimates for a set of mutually exclusive and exhaustive hypotheses exceeds 1.0 (referred to as subadditivity) (e.g. Robinson & Hastie, 1985; Teigen, 1974, 1983; Tversky & Koehler, 1994; Van Wallendael, 1989; Van Wallendael & Hastie, 1990; Windschitl, 2000; Wright & Whalley, 1983). One important aspect of likelihood judgment that has yet to be fully addressed by support theory concerns the processes by which people evaluate evidence for component hypotheses in the residual (i.e., the alternative hypotheses constituting the residual). Examples of possible processes include (a) analyzing individual components and then aggregating to assess overall strength, (b) sampling a portion of evidence from each component to estimate strength, and (c) making a global assessment of strength without an assessment of the strength for any individual hypothesis.

The heuristic comparison processes described here suggest that, when judging gut-level certainty, evidence assessment for the individual alternative hypotheses can be important, not just the overall assessment for the entire residual. It seems quite possible that the importance of individual alternative hypotheses could also apply to some of the types of subjective probability judgments addressed by support theory. If so, then the comparison heuristic may need to be integrated into explanations of some likelihood judgment phenomena. For example, one contributing cause to the subadditivity of subjective probability is that people attend to evidence supporting the focal outcome more than evidence supporting the alternatives (see, e.g., Tversky & Koehler, 1994; Sanbonmatsu, et al., 1997). However, this underweighting of the evidence for alternatives is unlikely to apply equally to all alternatives. Alternative-outcomes effects, particularly those detected in Experiment 3, indicate that alternative outcomes can differ dramatically in the attention or weight they receive when a focal outcome is being judged. Hence, the way in which evidence is distributed among a full set of outcomes may play an important role in the overall subadditivity of likelihood estimates for that set and in the degree to which individual outcomes are overestimated/underestimated.

Before concluding, we address a specific issue regarding the relationship between alternative-outcomes effects and the unpacking effects that are central to support theory. At first glance, these effects seem contradictory. Unpacking effects (e.g., see Fischhoff, Slovic, & Lichtenstein, 1978; Teigen, 1974; Tversky & Koehler, 1994) demonstrate that a packed hypothesis (e.g., that the next winner of the Boston Marathon will be from Africa) is judged as having less support than the summed support of its unpacked components (e.g., that the next winner would be from Kenya, Ethiopia, South Africa, or some other African nation). This finding would suggest that as an alternative (i.e., nonfocal) hypothesis is unpacked, the perceived support for that hypothesis would increase, resulting in a decrease in the judged likelihood of the focal hypothesis. However, in demonstrations of the alternative-outcomes effect, the judged likelihood of the focal hypothesis is lower when there is one strong "packed" alternative than when that strong alternative is "unpacked" (or, more accurately, replaced with weaker hypotheses that have the same overall level of objective support as the "packed" alternative). Perhaps the clearest example of this is the cookie jar study described earlier (Windschitl & Wells, 1998). The judged likelihood of drawing a chocolate-chip cookie was lower when the alternatives were seven oatmeal cookies than when they were different types of cookies that summed to seven. Hence, it seems that breaking an alternative hypothesis into component hypotheses can either decrease (as in unpacking effects) or increase (as in alternative-outcomes effects) the judged likelihood of the focal hypothesis.

How can this apparent contradiction be resolved? A key mediator in the typical unpacking demonstration is a memory/accessibility process. Unpacking an alternative hypothesis increases its overall judged probability (thus deflating the judged probability of the focal hypothesis) because the unpacked version reminds respondents of possibilities or specific outcomes (e.g., Kenyan runners and Ethiopian runners) that they would have overlooked or not recalled given only the packed hypothesis (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994). However, in demonstrations of alternative-outcomes effects, the roles of memory/accessibility are largely neutralized; the relevant possible outcomes are equally salient and accessible in the "packed" and "unpacked" versions. With memory processes neutralized, the heuristic comparison processes determine the direction of the alternative-outcomes effects. This explanation, which has yet to be directly tested, suggests that for some problems, breaking a nonfocal hypothesis into component hypotheses can have competing effects that are mediated by two distinct sets of processes: 1) the heuristic comparisons discussed in this article and 2) the mental recruitment of support for the nonfocal hypotheses.

Conclusion

While we have already drawn conclusions about the specific findings of this research, we end with a more general point. Historically, most research in the judgment and decision making domain has investigated people's "best" judgments and decisions. In other words, the participants are asked to think carefully and provide their best responses in the relevant tasks. This is true, even for most of the classic and current research on heuristics. The present experiments were an exception to this general rule. Our participants were told that we were interested in their gut reactions, not whether they could identify the objective probabilities of events. The results indicate that these gut-level responses were far from haphazard or aimless; after all, they did come from the head and not from the gut. The consistency of the data patterns within and across these three experiments provide important clues as to the organized set of processes that drive such gut-level judgments. As mentioned above, previous research on alternative-outcomes effects has already suggested that there are important differences between the processes that drive people's gut-level perceptions of certainty and those that drive their "best" judgments of probability (e.g., Teigen, in press; Windschitl & Wells, 1998). We suspect that there is much to be learned from future research that investigates gut-level perceptions of certainty in the context of other likelihood judgment phenomena. Such research would help provide a fuller understanding of the processes mediating both gut-level certainty judgments and "best" judgments of probability.

REFERENCES

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillside, NJ: Erlbaum.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology,* 66, 819–829.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault tress: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance,* 4, 330–344.

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

Gigerenzer, G., Todd, P. M., & the ABC Group (Eds.) (1999). *Simple heuristics that make us smart.* New York: Oxford Univ. Press.

Gonzalez, M., & Frenck-Mestre, C. (1993). Determinants of numerical versus verbal probabilities. *Acta Psychologica,* 83, 33–51.

Kirkpatrick, L. A., & Epstein, S. (1992). Cognitive-experiential self-theory and subjective probability: Further evidence for two conceptual systems. *Journal of Personality and Social Psychology,* 63, 534–544.

Robinson, L. B., & Hastie, R. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology*: *Human Perception and Performance,* 11, 443–456.

Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: Advances in support theory. *Psychological Review,* 104, 406–415.

Sanbonmatsu, D. M., Posavac, S. S., & Stasney, R. (1997). The subjective beliefs underlying probability overestimation. *Journal of Experimental Social Psychology,* 33, 276–295.

Teigen, K. H. (1974). Subjective sampling distributions and the additivity of estimates. *Scandinavian Journal of Psychology,* 15, 50–55.

Teigen, K. H. (1983). Studies in subjective probability III: The unimportance of alternatives. *Scandinavian Journal of Psychology,* 24, 97–105.

Teigen, K. H. (1988). When are low-probability events judged to be 'probable'? Effects of outcome-set characteristics on verbal probability estimates. *Acta Psychologica,* 67, 157–174.

Teigen, K. H. (in press). *When equal chances = good chances: Verbal probabilities and the equiprobability effect. Organizational Behavior and Human Decision Processes.*

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review,* 101, 547–567.

Van Wallendael, L. R. (1989). The quest for limits on noncomplementarity in opinion revision. *Organizational Behavior and Human Decision Processes,* 43, 385–405.

Van Wallendael, L. R., & Hastie, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory and Cognition,* 18, 240–250.

Windschitl, P. D. (2000). The binary additivity of subjective probability does not indicate the binary complementarity of perceived certainty. *Organizational Behavior and Human Decision Processes,* 81, 195–225.

Windschitl, P. D., & Weber, E. U. (1999). The interpretation of "likely" depends on the context, but "70%" is 70%—right?: The influence of associative processes on perceived certainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 25, 1514–1533.

Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied,* 2, 343–364.

Windschitl, P. D., & Wells, G. L. (1998). The alternative-outcomes effect. *Journal of Personality and Social Psychology,* 75, 1411–1423.

Wright, G., & Walley, P. (1983). The supra-additivity of subjective probability. In B. P. Stigum & F. Wenstøp (Eds.), *Foundations of utility and risk theory with applications* (pp. 233–244). Dordrecht, The Netherlands: Reidel.

Received February 2, 2000; published online April 5, 2001