

Likelihood judgment based on previously observed outcomes: The alternative-outcomes effect in a learning paradigm

PAUL D. WINDSCHITL
University of Iowa, Iowa City, Iowa

MICHAEL E. YOUNG
Southern Illinois University, Carbondale, Illinois

and

MARY E. JENSON
University of Iowa, Iowa City, Iowa

Previous research has demonstrated that intuitive perceptions of certainty regarding a focal outcome are sensitive to variations in how evidence supporting nonfocal alternatives is distributed, even when such variations have no bearing on objective probability. We investigated this *alternative-outcomes effect* in a learning paradigm in which participants made likelihood judgments on the basis of their memory for past observations of relevant outcomes. In Experiment 1, a manipulation of evidence (observed frequencies) across alternative outcomes influenced not only intuitive certainty estimates about a focal outcome but also numeric subjective probabilities. Experiment 2 ruled out the possibility that these effects were attributable to the influence of information loss on frequency estimations. The findings were consistent with the heuristic comparison account, which suggests that the judged likelihood of a focal outcome will be disproportionately influenced by the strength (frequency) of the strongest alternative outcome.

When people judge the likelihood of a particular outcome to an event, how do possible alternative outcomes influence certainty? As the overall strength of evidence for the alternatives increases, certainty in the focal outcome should decrease. There is a wealth of research relevant to determining the extent to which people's certainty judgments conform to this rule (e.g., Sanbonmatsu, Posavac, & Stasney, 1997; Teigen, 1974, 1983; Tversky & Koehler, 1994; Van Wallendael & Hastie, 1990; Windschitl, 2000; Wright & Whalley, 1983). The present research concerns a slightly different question: Is certainty in a focal outcome sensitive to variations in the *distribution* of evidence across alternative outcomes? Recent studies have revealed that people's perceptions of certainty about a focal outcome are often 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, 2001; Windschitl & Wells, 1998; Windschitl & Young, 2001). Participants in one study, for example, expressed greater

optimism about winning a hypothetical 59-ticket raffle when they held 17 tickets and others held 9, 9, 8, 8, and 8 versus when they held 17 and others held 16, 7, 7, 6, and 6 (Windschitl & Young, 2001). In another study, participants reading about a young girl indicated greater certainty that she would draw a chocolate chip cookie when the cookie jar she was drawing from was said to contain 2 chocolate chip cookies, 1 oatmeal, 1 raisin, 1 butterscotch, 1 rum, 1 peanut butter, 1 pecan, and 1 sugar cookie, than when the jar was said to contain 2 chocolate chip cookies and 7 oatmeal cookies (Windschitl & Wells, 1998).

Windschitl and Wells (1998) referred to the effect that these types of manipulations can have on perceptions of certainty as an *alternative-outcomes effect*, and they argued that the effect provides insights about the processes by which people consider evidence for alternative outcomes when judging a focal outcome's likelihood. Specifically, they argued that the alternative-outcomes effect reflects the influence that a heuristic comparison process has on intuitive perceptions of certainty (see also Windschitl & Young, 2001). When judging certainty of a focal outcome, people appear to compare the strength of the focal outcome with the strength of the strongest alternative outcome. The more this comparison favors the focal outcome—or the less it favors the strongest alternative—the

This work was supported by Grant SES 99-11245 from the National Science Foundation. Correspondence should be addressed to P. D. Windschitl, Department of Psychology, University of Iowa, Iowa City, IA 52242 (e-mail: paul-windschitl@uiowa.edu).

greater the perceived likelihood of the focal outcome (for supporting evidence, see Windschitl & Young, 2001). This comparison can serve as a useful heuristic that provides a roughly accurate guide as to how optimistic or pessimistic one should feel about the possibility of the focal outcome. However, it can sometimes produce differences in perceived certainty (e.g., alternative-outcomes effects) that depart from normative standards.

Like previous research on the alternative-outcomes effect, the present studies focused on how people judge the certainty of a focal outcome when there are multiple alternatives to that outcome. However, unlike previous research, which utilized a judgment paradigm in which participants were given explicit summaries of outcome frequency, the present project examined likelihood judgments that were based on direct observations of past outcomes. Specifically, participants viewed a series of stimuli that represented events with varying outcomes. The frequencies of the outcomes were systematically varied to create an "alternative-outcomes manipulation" similar to those used in the research described above (e.g., Windschitl & Wells, 1998). In other words, while the frequencies of individual nonfocal outcomes differed depending on the manipulation, the sum of the frequencies across the set of outcomes was held constant. After observing the outcomes, participants were asked about the likelihood that a particular outcome would be selected in a random draw from a set of possible outcomes.

In employing this learning paradigm, we hoped to achieve two main goals. One goal was to establish evidence either for or against the notion that the alternative-outcomes effect and its mediating processes operate broadly on all categories of likelihood judgments, not simply those tested within the previously described judgment paradigm. Although we expected to detect alternative-outcomes effects in our learning paradigm, there was also a plausible rationale for expecting the absence of such effects. Consider a case in which participants are asked to judge the likelihood that Outcome A will occur rather than B, C, or D. In a judgment paradigm, the evidence relevant for this judgment would be presented in a structured format; the frequencies of A, B, C, and D would be summarized as four separate values. This structure might facilitate heuristic comparisons between the focal and other *individual* outcomes. In a learning paradigm, however, outcome instances are observed in a random order. Given this lack of structure, a respondent who is judging the likelihood of A may forgo representing B, C, and D separately. Instead, he/she may search memory for instances of A and then search memory for instances that fit into an aggregate category of "all other outcomes." In other words, participants may treat a nonbinary question (How likely is A rather than B or C or D?) as a binary one (How likely is A rather than not-A?). If so, there is no reason to expect that a heuristic comparison between the focal outcome and the strongest individual alternative would produce an alternative-outcomes effect. Hence, detecting alternative-outcomes

effects in the present experiments would indicate that even when evidence for possible outcomes is encountered in an unorganized fashion, internal representations of evidence at the level of individual nonfocal outcomes—not just an aggregate level—are critical to judging the likelihood of a focal outcome.

A second goal related to the use of the learning paradigm was to determine whether the processes mediating alternative-outcomes effects influence numeric as well as verbal expressions of certainty. In previous research, alternative-outcomes manipulations have tended to produce robust effects on verbal and other nonnumeric measures (e.g., a measure that includes response options like "somewhat likely") but no discernible effects on numeric probability measures (e.g., Teigen, 1988, 2001; Windschitl & Wells, 1996, 1998; Windschitl & Young, 2001). This asymmetry may seem to suggest that alternative-outcomes effects are unique to vague uncertainty measures and/or to people's nondeliberative responses. We predicted, however, that a numeric probability measure would detect alternative-outcomes effects in a learning paradigm. We assumed that heuristic comparison processes have the potential to produce alternative-outcomes effects whenever a respondent cannot use a purely rule-based strategy for calculating or mapping from the problem input (e.g., frequency information) to a specific response. In a judgment paradigm, participants providing numeric probabilities (but not those providing verbal estimates) could use a purely rule-based strategy because the necessary numbers for computations are readily available. In a learning paradigm, however, participants would not know the exact frequencies of the previously observed outcomes, and therefore would be unlikely to attempt to employ a purely rule-based calculation strategy for determining numeric probability. Instead, they would either ignore the calculation strategy in favor of other estimation strategies or they would apply some sort of estimated calculation strategy—just as one might do when roughly estimating the answer to a long division problem without working out all of the relevant computations. We believe that these estimation strategies allow heuristic processes like those that mediate alternative-outcomes effects to influence final probability responses.

EXPERIMENT 1

Experiment 1 was designed to test for an alternative-outcomes effect in a learning paradigm. According to the heuristic comparison hypothesis and findings from previous research involving a judgment paradigm (Windschitl & Young, 2001), the size of an alternative-outcomes effect depends primarily on the extent to which the relevant manipulation varies the strength of the *strongest* alternative outcome; a manipulation will have the greatest impact on the perceived certainty of a focal outcome when the strongest alternative outcome is weak in one condition but quite strong in the other condition. Hence, for the present

experiment we used a manipulation in which a focal outcome was observed 30 times, but the possible alternatives to the focal outcome were observed 45, 5, 5, and 5 times or 15, 15, 15, and 15 times. After observing outcomes, participants indicated on both verbal and numeric scales their subjective certainty that, in a random drawing from the observed outcomes within a distribution, the focal outcome would happen to be selected. Consistent with the heuristic comparison hypothesis, we expected that both verbal and numeric responses would indicate significantly greater certainty about a focal outcome that was part of the 30-15-15-15-15 distribution than a focal outcome that was part of the 30-45-5-5-5 distribution, even though the focal outcome would have an objective probability of .33 in both cases. Before completing the experiment, participants also provided additional numeric probability estimates as well as frequency estimates for all outcomes.

Method

Participants and Design. The participants were 40 students enrolled in introductory psychology courses at the University of Iowa. The alternative-outcomes manipulation was implemented within participants, and counterbalancing manipulations were between participants.

Stimuli. Twelve computer-presented photographs were used to represent events and outcomes. The beginning of each event was represented by an image of a postal package with either a toolbox or feet drawn on it. Five labeled images represented a set of tool outcomes (a screwdriver, utility knife, wrench, pliers, and tape measure), and five labeled images represented a set of footwear outcomes (canvas, hiking, dress, running, and brown shoes).

Procedure. Groups of 1-6 participants were asked to imagine that they had a friend who sent gifts—tools and footwear. They were also asked to imagine that they would always open the packages they received but would then repack them and throw them onto either a tool pile or a footwear pile in their garage. After reading about this setup story, participants read the following: "So, in the first part of this experiment, you will see a series of packages, each with a picture of a toolbox or feet. The package will then open, so you will see what item is inside. After seeing what the item is, press the key that corresponds to the first letter of the name of the item." The computer program then began showing a series of "events" with "outcomes." Each event started with a package (1 sec) that was immediately followed by a tool or footwear outcome. When the appropriate key was pressed, another event-outcome sequence was immediately presented.

Each participant saw 180 event-outcome sequences in a random order; 90 involved tools and 90 involved footwear. The focal outcome from each set always appeared 30 times. For category counterbalancing, participants saw either a 30-45-5-5-5 distribution for tools and a 30-15-15-15-15 distribution for footwear, or vice versa. We also ensured that each outcome served as the focal one equally often across participants and the two distributions.

After seeing these event-outcome sequences, participants were reminded in a paper questionnaire that their garage now contained one unorganized pile of tool packages and a second pile of footwear packages. They were asked to imagine that they had a particular need for the focal tool and had gone to the tool pile to randomly pick a package. They indicated on a verbal certainty scale how likely it seemed that they would happen to select the specified tool. This scale contained 11 verbal-certainty expressions (e.g., "rather likely") evenly spaced above a line, and participants responded by placing a mark across that line. A similar question was then posed for the

footwear. Next, participants were asked to estimate the numeric probability of choosing the specified tool and footwear (e.g., "What is the chance that when looking for a screwdriver, you would happen to pick a box with a screwdriver in it?"). A brief description of how to use the probability scale (actually a 0% to 100% scale) was included. Finally, participants provided probability estimates for the remaining possible outcomes, followed by absolute frequency estimates for each of the outcomes.

Results and Discussion

Certainty judgments. The responses on the verbal certainty scale were scored from 0 (*impossible*) to 100 (*certain*). The verbal and numeric estimates for the focal outcomes were submitted to a 2 (verbal and numeric response type) \times 2 (distribution type) \times 2 (category counterbalance) mixed model analysis of variance (ANOVA) with response type and distribution type as repeated measures. The response-type main effect was significant, but because the values from the verbal and numeric scales cannot be considered directly comparable, this main effect will not be discussed further (see Windschitl & Wells, 1996). The main effect for the counterbalancing factor was not significant, and no interactions were significant (all p s > .20). The key finding from the ANOVA was the predicted significant main effect for distribution type [$F(1,38) = 4.50, p < .05$]. Consistent with the direction of alternative-outcomes effects detected within judgment paradigms, participants provided higher certainty estimates for the focal outcome in the 30-15-15-15-15 distribution than in the 30-45-5-5-5 distribution. The respective means on the verbal certainty measures were 66.8 ($SD = 18.2$) and 60.7 ($SD = 19.1$); the respective means on the numeric measures were 52.4 ($SD = 23.0$) and 44.6 ($SD = 22.4$).

Table 1 displays the mean subjective probabilities for all outcomes (verbal estimates were collected only for the focal outcomes). This table shows that participants tended to

Table 1
Probability Estimates for Focal and Alternative Outcomes
as a Function of Outcome Frequency in Experiment 1

| Distribution Type Frequency of Presentation | Objective Probability | Mean Probability Estimate | Standard Deviation |
|--|--------------------------|---------------------------------|-----------------------|
| 30-15-15-15-15 distribution | | | |
| 30 (focal) | 33.3 | 52.4 | 23.0 |
| 15 | 16.6 | 25.0 | 13.6 |
| 15 | 16.6 | 27.6 | 20.8 |
| 15 | 16.6 | 27.9 | 18.3 |
| 15 | 16.6 | 23.3 | 14.3 |
| 30-45-5-5-5 distribution | | | |
| 30 (focal) | 33.3 | 44.6 | 22.4 |
| 45 | 50.0 | 57.8 | 24.8 |
| 5 | 5.5 | 13.6 | 11.1 |
| 5 | 5.5 | 12.7 | 9.0 |
| 5 | 5.5 | 11.3 | 7.8 |

Note—The values in the table collapse across counterbalancing schemes and specific category items. For example, the first value in the Mean Probability Estimate column is the average subjective probability estimate given for whatever item was seen 30 times as part of the 30-15-15-15-15 distribution. All probability values are displayed as percentages.

overestimate the numeric probability of the outcomes—a finding that is often observed in likelihood-judgment studies (e.g., Sanbonmatsu et al., 1997; Teigen, 1974, 1983; Tversky & Koehler, 1994; Windschitl, 2000). Despite this overestimation and resulting nonadditivity, participants' probability estimates were generally quite sensitive to the objective probabilities. We computed a correlation for each participant between his/her probability judgments and the relevant objective probabilities. The mean of these correlations was .83. Consequently, it seems fair to say that the nonnormative alternative-outcomes effects that were detected were produced in the context of a largely coherent pattern of certainty responses.

The observed alternative-outcomes effect suggests that participants' judgments cannot be easily simplified and conceptualized as judgments about a binary case (How likely will the focal outcome occur versus not occur?). At some point in the judgment process, participants were sensitive to how the frequencies of alternative outcomes were distributed across those individual outcomes. Similar findings from a judgment paradigm have been attributed to a heuristic comparison between the focal outcome and the strongest of the alternative outcomes, and we suspect the same process operated here. For example, when judging the likelihood of a hammer, participants did not merely compare the frequency of the hammer with an aggregated frequency of all tools. Instead, or perhaps in addition, their likelihood judgment was shaped by a comparison between the subjective frequencies of the hammer and the strongest alternative outcome.

The results from the ANOVA suggest that alternative-outcomes effects are not linked to some unique property of vague certainty measures; they extend to numeric measures as well. This conclusion, however, is tempered by a particular aspect of the procedures used in Experiment 1. Specifically, all participants provided verbal certainty estimates about the focal outcomes immediately before they provided numeric estimates. Perhaps the verbal measure prompted processes that were sensitive to the alternative-outcomes manipulation, and these processes or the resulting responses on the verbal scales had carryover effects on the numeric responses. Experiment 2 more effectively resolved the question of whether numeric measures exhibit alternative-outcomes effects when they are not preceded by other measures.

Frequency judgments. The means of the subjective frequencies for each outcome are displayed in Table 2. This table shows that participants were, at a group level, sensitive to the actual frequencies of the outcomes. To examine this sensitivity at the level of individuals, we computed a correlation for each participant between his/her frequency judgments and the relevant objective frequencies. We then computed the mean of these correlations across participants; we conducted appropriate Fisher z conversions and reconversions when calculating this mean and other means of the same type reported elsewhere in this paper. The resulting mean correlation was quite high ($r = .87$).

Despite the orderly nature of participants' frequency estimates, an analysis of their estimates for the focal outcomes, both seen 30 times, revealed a surprising finding. Participants' frequency estimates were significantly higher for the focal outcome in the 30-15-15-15-15 distribution ($M = 27.3$) than in the 30-45-5-5-5 distribution ($M = 23.5$) [$F(1,38) = 4.71, p < .05$]. This effect, which was not qualified by an interaction with the category-counterbalancing factor ($F < 1$), suggests a different possible explanation for the alternative-outcomes effects detected on the certainty measures. According to that explanation, the alternative-outcomes manipulation influenced judgments of certainty because it changed the subjective frequencies of the focal outcomes themselves, not because of heuristic processes unique to likelihood judgment. This explanation, however, received a more thorough empirical test in Experiment 2. An additional test was required because the frequency measures in Experiment 1 always followed the certainty measures, which raised the possibility that the effect detected on the frequency measures could be attributable to some form of carryover effect prompted by the certainty measures.

The frequency responses also showed a characteristic regression effect: High-frequency outcomes (those observed 30 or 45 times) were underestimated, low-frequency outcomes (observed 5 times) were overestimated, and medium-frequency outcomes (observed 15 times) were estimated with little systematic bias. As discussed by Fiedler and Armbruster (1994), this type of regression effect can be understood in terms of the intrusion of error variance into the processing or storage of frequency information. Participants must learn deviations from equality in the frequencies of the outcomes. To the extent that this process is imperfect—because of encoding deficiencies, memory loss, or other forms of information loss—the learned or subjective frequencies would be regressive. That is, the

Table 2
Frequency Estimates for Focal and Alternative Outcomes
as a Function of Outcome Frequency in Experiment 1

| Distribution Type Frequency of Presentation | Mean Frequency Estimate | Standard Deviation |
|--|-------------------------------|-----------------------|
| 30-15-15-15-15 distribution | | |
| 30 (focal) | 27.3 | 11.9 |
| 15 | 15.5 | 8.5 |
| 15 | 14.8 | 8.6 |
| 15 | 16.1 | 10.6 |
| 15 | 15.1 | 8.5 |
| 30-45-5-5-5 distribution | | |
| 30 (focal) | 23.5 | 10.7 |
| 45 | 31.1 | 12.2 |
| 5 | 8.2 | 5.8 |
| 5 | 7.8 | 4.9 |
| 5 | 7.6 | 6.2 |

Note—The values in the table collapse across counterbalancing schemes and specific category items. For example, the first value in the Mean Frequency Estimate column is the average subjective frequency estimate given for whatever item was seen 30 times as part of the 30-15-15-15-15 distribution.

distribution of subjective frequencies across outcomes would be flatter than the distribution of actual frequencies (see, e.g., Attneave, 1953; Fiedler & Armbruster, 1994; Hintzman, 1988; Sedlmeier, Hertwig, & Gigerenzer, 1998).

Although this type of regression effect in frequency judgments has been well documented, the ultimate impact of this regression effect on probability judgments in non-binary cases has not received attention. In the present work, regression effects on subjective frequency estimates have the potential to influence focal-outcome probability judgments in a direction opposite that of a heuristic comparison process. Consider specifically the subjective frequencies of outcomes in the 30-45-5-5-5 distribution. If the regression-related deflation of the highest frequency outcome is greater than the overall inflation of the low-frequency outcomes, then the sum of the subjective frequencies for nonfocal outcomes might underestimate the objective value (60), which could potentially inflate the judged certainty of the focal outcome. Indeed, the mean subjective frequencies for the nonfocal alternatives in the 30-45-5-5-5 distribution were 31.1, 8.2, 7.8, and 7.6, which yields a mean sum of only 54.7. This sum was significantly less than the analogous sum for the 30-15-15-15-15 distribution (61.5) [$F(1,38) = 5.12, p < .05$].

To summarize, participants' subjective frequencies for the focal outcomes were higher in the 30-15-15-15-15 distribution than in the 30-45-5-5-5 distribution, which was perhaps due to a carryover effect. At the same time, their subjective frequency estimates for the nonfocal alternatives were also higher in the former rather than latter distribution, which was likely due to regression effects. Not surprisingly, then, when we conducted an analysis of what we call "externally calculated" probability estimates—which were created by taking each participant's own frequency estimate for a focal outcome and dividing by the sum of his/her frequency estimates across all possible outcomes in a set—these externally calculated estimates showed no significant alternative-outcomes effect [$F(1,38) < 1$]. The mean externally calculated probability of the focal outcome was 31.7 ($SD = 7.5$) for the 30-15-15-15-15 distribution and 30.1 ($SD = 6.8$) for the 30-45-5-5-5 distribution.

EXPERIMENT 2

Experiment 1 provided clear evidence of an alternative-outcomes effect in a learning paradigm, but it was ambiguous regarding two types of mediational accounts for the effect. Both types of accounts assume that people use subjective representations of frequency when judging focal-outcome certainty. The first type suggests that the alternative-outcomes manipulation biases the estimated frequencies of the focal outcomes. The second type suggests that frequency estimates, per se, are not systematically biased by the manipulation. Instead, the way in which frequency information is used to make certainty judgments makes those certainty judgments sensitive to the manipulation.

Regarding the first type of account, there is a broad range of theories about how people track and judge the frequency

of objects, words, events, and other entities (e.g., Betsch, Siebler, Marz, Hormuth, & Dickenberger, 1999; Brown, 1995, 1997; Greene, 1986, 1989; Hasher & Zacks, 1984; Hintzman, 1988, 2001; Howell, 1973; Jonides & Naveh-Benjamin, 1987; Maki & Ostby, 1987; Manis, Shedler, Jonides, & Nelson, 1993; Sedlmeier et al., 1998). Although these theories can account for a variety of frequency judgment effects, including some context effects, they do not provide specific predictions about how the judged frequency of a particular object would be influenced by the distribution of frequencies across individual objects in the same context or category. Nevertheless, one could propose extensions of existing models to account for such potential influence. For example, models that assume that frequency judgment is mediated by a recall process might also assume that when participants are judging the frequency of a focal outcome in a 30-45-5-5-5 distribution, recall of the item seen 45 times is so pervasive that it reduces recall attempts for the focal outcome, thereby lowering the judged frequency of that outcome. Models that assume that associative strengths underlie frequency judgments might also assume that even when an absolute judgment is solicited, the *relative* difference between the associative strengths of the focal outcome and the strongest alternative outcome plays a mediating role.

Although a frequency-bias explanation is fully consistent with the results of Experiment 1, the ultimate tenability of this type of explanation hinges on whether the significant effects found on focal-outcome frequency estimates were products of carryover effects—a possibility that we suggested earlier. Therefore, we designed a second experiment that closely resembled Experiment 1, but we manipulated whether participants made frequency judgments before probability judgments or vice versa. If focal-outcome frequency judgments exhibit sensitivity to the alternative-outcomes manipulation even when they are not preceded by certainty measures, then the alternative-outcomes effects on certainty measures in this paradigm were likely mediated by frequency-estimation biases. However, if those same focal-outcome frequency judgments show little sensitivity to the manipulation, this would not bode well for the frequency-estimation-bias hypothesis. Instead, it would appear that likelihood judgment processes, such as the comparison heuristic, would provide the better account for the alternative-outcomes effects.

A second important feature of Experiment 2 was that verbal certainty measures always followed rather than preceded the questions about the probabilities and frequencies of the focal outcomes. This feature allowed us to address the question of whether the significant alternative-outcomes effect detected with the numeric probability measure in Experiment 1 was due to a carryover effect involving the preceding verbal certainty measure.

Method

Participants and Design. The participants were 200 students enrolled in introductory psychology courses at the University of Iowa. The alternative-outcomes manipulation was implemented within

Table 3
Probability Estimates from Participants in the
Probability-First Group of Experiment 2

| Distribution Type Frequency of Presentation | Mean | | Standard Deviation |
|--|--------------------------|-------------------------|-----------------------|
| | Objective Probability | Probability Estimate | |
| <i>18-9-9-9-9</i> distribution | | | |
| 18 (focal) | 33.3 | 41.7 | 18.3 |
| 9 | 16.6 | 22.5 | 12.8 |
| 9 | 16.6 | 23.7 | 13.5 |
| 9 | 16.6 | 23.1 | 13.2 |
| 9 | 16.6 | 21.7 | 12.3 |
| <i>18-27-3-3-3</i> distribution | | | |
| 18 (focal) | 33.3 | 37.9 | 16.9 |
| 27 | 50.0 | 47.2 | 21.1 |
| 3 | 5.5 | 12.0 | 10.7 |
| 3 | 5.5 | 13.5 | 10.1 |
| 3 | 5.5 | 12.6 | 11.0 |

Note—The values in the table collapse across counterbalancing schemes and specific category items. For example, the first value in the Mean Probability Estimate column is the average subjective probability estimate given for whatever item was seen 18 times as part of the 18-9-9-9-9 distribution. All probability values are displayed as percentages.

participants. Counterbalancing manipulations and a probability-first versus frequency-first manipulation were implemented between participants.

Procedures and Materials. The procedures and stimuli were identical to those of Experiment 1 with the following three exceptions. First, the presentation frequencies of all outcomes were reduced by 40%. Hence, instead of observing distributions of 30-15-15-15-15 and 30-45-5-5-5, participants observed distributions of 18-9-9-9-9 and 18-27-3-3-3. This change was introduced to reduce the length of the outcome presentation, which some participants in Experiment 1 found to be overly monotonous. Second, participants in the probability-first group answered the dependent measures in the following order: probability questions about the two focal outcomes and then about the alternative outcomes, frequency questions about the two focal outcomes and then about the alternative outcomes, and finally verbal certainty questions about the two focal outcomes. Participants in the frequency-first group responded to the frequency questions, the probability questions, and finally the verbal certainty questions. Third, the verbal certainty questions were slightly changed in order to more explicitly emphasize to the participants that their responses should reflect their intuitive or "gut-level" impressions. Although we believe that verbal measures are generally successful at encouraging some degree of intuitive responding, we wanted to augment this characteristic because the verbal measures in Experiment 2 were preceded by numeric certainty and frequency measures, which might tend to prompt more rule-based responding. The verbal questions took the following form: "Now . . . we are interested in how you *feel* about whether a specific outcome will occur. In other words, we are interested in your gut-level reaction, not your well-thought-out or calculated response. Imagine again that you pick a box from the pile of tools and hope it contains a screwdriver. How good do you feel about the chance you happened to pick a box that contains a screwdriver?" (1 = *not good at all*; 9 = *very good*).

Results and Discussion

Probability judgments. Because of our concern regarding potential carryover effects, our primary analyses focused on the data that participants provided first—the probability judgments from the probability-first group and the frequency judgments from the frequency-first group. The initial question of interest was whether the

probability judgments from the probability-first group exhibited alternative-outcomes effects, even though those judgments were not preceded by verbal certainty judgments. As expected, a 2 (distribution type) \times 2 (category counterbalance) ANOVA on their probability judgments revealed a significant alternative-outcomes effect [$F(1,98) = 4.94, p < .05$]. Probability judgments were higher for the focal outcome in the 18-9-9-9-9 distribution ($M = 41.7, SD = 18.3$) than in the 18-27-3-3-3 distribution ($M = 37.9, SD = 16.9$). The effects of the counterbalancing factor and interaction were not significant ($F_s < 1$).

The magnitude of the alternative-outcomes effect was somewhat smaller in Experiment 2 than in Experiment 1. Also, Table 3, which displays the mean responses from the probability-first group, shows that the probability judgments for most outcomes tended to be smaller in Experiment 2 than in Experiment 1. Determining the cause of these differences is difficult because there are so many causal candidates. One possible cause is the fact that the subjective probability measures in Experiment 1 were preceded by verbal certainty measures, which may have facilitated more heuristic processing and reduced participants' concern about keeping probability estimates small enough to remain additive (see, e.g., Windschitl, 2000). Other possible causes may be related to the fact that the frequencies of the presented outcomes in Experiment 2 were reduced by 40%. Regardless of how the differences between Experiments 1 and 2 are explained, the key finding is that numeric probability judgments, even when not preceded by verbal certainty measures, are sensitive to the distribution of evidence across alternative outcomes.

Frequency judgments. The second critical set of analyses concerned frequency judgments of participants in the frequency-first group (Table 4). A 2 (distribution type) \times 2 (category counterbalance) ANOVA on the subjective frequencies for the focal outcomes revealed nonsignifi-

Table 4
Frequency Estimates for Participants in the
Frequency-First Group in Experiment 2

| Distribution Type Frequency of Presentation | Mean | Standard Deviation |
|--|-----------------------|-----------------------|
| | Frequency Estimate | |
| <i>18-9-9-9-9</i> distribution | | |
| 18 (focal) | 17.4 | 11.2 |
| 9 | 11.6 | 6.9 |
| 9 | 12.0 | 6.7 |
| 9 | 11.5 | 6.6 |
| 9 | 11.5 | 6.9 |
| <i>18-27-3-3-3</i> distribution | | |
| 18 (focal) | 17.6 | 11.1 |
| 27 | 24.9 | 11.3 |
| 3 | 4.8 | 3.5 |
| 3 | 5.6 | 4.4 |
| 3 | 4.9 | 3.0 |

Note—The values in the table collapse across counterbalancing schemes and specific category items. For example, the first value in the Mean Frequency Estimate column is the average subjective frequency estimate given for whatever item was seen 18 times as part of the 18-9-9-9-9 distribution.

cant effects for distribution type [$F(1,98) < 1$], the counterbalancing factor [$F(1,98) < 1$], and the interaction [$F(1,98) = 1.47, p > .05$]. Not only was there no significant effect for the alternative-outcomes manipulation, but the trend in the data was also opposite to that found in Experiment 1. Hence, it seems likely that the frequency judgment effect detected in Experiment 1 was due to carryover processes or a Type I error.

As in Experiment 1, the correlation between a participant's frequency judgment and the true frequency tended to be quite strong ($M_r = .87$). Also as in Experiment 1, the participants' frequency judgments exhibited regression effects. As discussed earlier, these regression effects have the potential to produce notable effects on externally calculated probabilities. Indeed, when we used the participants' own frequency estimates to calculate probability values for the frequency-first group, the calculated values showed a significant effect in the direction *opposite* that of the typical alternative-outcomes effect [$F(1,98) = 6.93, p < .05$]. The mean externally calculated probability of the focal outcome was 27.0 ($SD = 8.0$) for the 18-9-9-9-9 distribution and 29.4 ($SD = 8.7$) for the 18-27-3-3-3 distribution. This finding suggests that the key mechanisms producing alternative-outcomes effects in certainty judgments are not located in a frequency learning or estimation stage. Information loss leading to regression effects in the estimation of frequencies appears to operate in a direction that is opposite to that of the typical alternative-outcomes effect.

"Second" judgments. We also analyzed the second sets of judgments that participants provided—that is, the frequency judgments of the probability-first group and the probability judgments of the frequency-first group. Among the probability-first participants, the alternative-outcomes manipulation had no significant influence on focal-outcome frequency estimates [$F(1,98) = 1.5, p > .10$]. The means for the 18-9-9-9-9 and 18-27-3-3-3 distributions were 17.9 ($SD = 9.5$) and 17.0 ($SD = 8.9$), respectively. This null effect, in conjunction with the null effect found among the frequency-first participants in this experiment, suggests that the significant effect found on the frequency measure in Experiment 1 is likely attributable to a Type I error.

Regarding the probability judgments made by frequency-first participants, no significant alternative-outcomes effect was detected [$F(1,98) < 1$]. The means for the 18-9-9-9-9 and 18-27-3-3-3 distributions were 33.4 ($SD = 18.4$) and 32.5 ($SD = 19.9$), respectively. To explicitly test the influence of order (either frequency or probability first) on probability judgment, we conducted an ANOVA that included order, distribution, and counterbalancing as factors. Although the order \times distribution interaction was not significant [$F(1,196) = 1.1, p > .20$], the order main effect was significant [$F(1,196) = 9.5, p < .01$]. The frequency-first group gave lower and more accurate focal-probability judgments ($M = 33.0$) than did the probability-first group ($M = 39.8$), suggesting that probability judgments were influenced by carryover effects prompted by preceding frequency questions.

Verbal certainty judgments. Although verbal certainty judgments were solicited after frequency and probability judgments, we suspected that an alternative-outcomes effect might nevertheless emerge for the verbal judgments about focal outcomes, given that participants were encouraged to provide their intuitive or gut-level responses. Indeed, a 2 (distribution type) \times 2 (category counterbalance) \times 2 (probability or frequency first) ANOVA revealed no significant interactions or main effects, except for a significant alternative-outcomes effect [$F(1,196) = 9.74, p < .01$]. Participants expressed more certainty about the focal outcome in the 18-9-9-9-9 distribution ($M = 5.9, SD = 1.6$) than in the 18-27-3-3-3 distribution ($M = 5.5, SD = 1.7$).

GENERAL DISCUSSION

The conclusions supported by our experiments can be grouped around two related findings—the detection of alternative-outcomes effects on a numeric probability measure and the detection of alternative-outcomes effects within a learning paradigm.

Alternative-Outcomes Effects and Numeric Probability

As mentioned earlier, alternative-outcomes manipulations in past research have not had significant effects on participants' numeric probability estimates, even when they did have a systematic effect on verbal certainty estimates (e.g., Teigen, 2001; Windschitl & Wells, 1998). However, consistent with our hypothesis regarding the relationship between the alternative-outcomes effect and methods of likelihood measurement, reliable alternative-outcomes effects were detected on numeric measures in our experiments. Rather than assuming that some likelihood measures are sensitive to the effects and others are not, we suggest that a given likelihood measure will be insensitive to alternative-outcomes effects when a respondent engages an exclusively rule-based strategy for calculating or mapping from the problem input (e.g., frequency information about all outcomes) to a specific response (e.g., a numeric probability estimate). In a judgment paradigm, a respondent who is asked to provide a numeric probability estimate can successfully use this exclusively rule-based strategy. However, in this learning paradigm, respondents who were asked to provide a numeric probability estimate were likely to find it difficult to employ a purely rule-based strategy because they did not have precisely quantified knowledge of the relevant frequencies that would serve as the input in a rule-based calculation of probability. We assume that participants' responses involved some degree of estimation, not pure calculation. This estimation component was affected by the manner in which evidence was distributed across alternative outcomes. In principle then, the alternative-outcomes effect and the processes that mediate it may be as applicable to numeric subjective probability judgments as to verbal certainty judgments.

Alternative-Outcomes Effects in a Learning Paradigm

Experiments 1 and 2 demonstrated that the alternative-outcomes effect is not a phenomenon that is confined to a judgment paradigm. The results provide initial support for the position that the processes mediating the alternative-outcomes effect apply broadly to the various types of cases in which likelihood judgments are made. In the introduction, we noted a plausible rationale as to why likelihood judgments in a learning paradigm might not exhibit alternative-outcomes effects. It seemed possible that when people make judgments on the basis of outcomes viewed in an unorganized order—rather than discretely summarized frequency information—they would conduct a memory search in which the possible nonfocal outcomes for an event would be treated as an aggregated whole. However, our results indicate that representations of evidence at the level of *individual* outcomes—not just an aggregate level—are critical when participants are judging the likelihood of a focal outcome, even when the evidence for the alternatives must be organized and estimated by the respondents themselves. In other words, a likelihood judgment about a focal outcome with multiple alternatives cannot be easily simplified and conceptualized as a judgment about a binary case in which a respondent considers only two possibilities—either the focal outcome will or will not occur.

The heuristic comparison account that has been applied to alternative-outcomes effects in judgment paradigms provides a plausible explanation of the effects detected here. According to this account, comparisons between evidence for the focal outcome and evidence for individual alternatives—particularly the strongest alternative—play a key role in people's perceptions of certainty. Consistent with this account, Windschitl and Young (2001) demonstrated that the evidence for the strongest of the alternative outcomes has a far greater (and disproportionate) influence on the judged certainty of the focal outcome than does the second strongest alternative or other weaker alternatives. Relatedly, a study by Windschitl and Wells (1998) demonstrated that, relative to other nonfocal alternatives, participants were most interested in learning more information about the strength of the strongest alternative outcome when thinking about the possibility of a focal outcome (their winning) in a raffle. Hence, we suspect that a key causal component of the alternative-outcomes effects detected here was participants' generalized tendency to give the strongest alternative outcome disproportionate weight when judging a focal outcome.

Two additional explanations of the observed alternative-outcomes effects were ruled out. First, a significant effect on the focal-outcome frequency estimations in Experiment 1 raised the possibility that the alternative-outcomes manipulation influenced the learning or estimation of frequencies for the focal outcomes. However, the frequency estimates from both the probability-first and frequency-first groups in Experiment 2 suggest that this effect is best attributed to a carryover effect or Type I error.

A second possible explanation was that biases in the estimated frequencies of nonfocal outcomes caused the alternative-outcomes effects. However, our results indicated that regression effects in these experiments operated in a direction that was opposite that of the alternative-outcomes effect detected on certainty measures. More specifically, when we calculated probability values from participants' own frequency estimates in Experiment 2, those values for the two focal outcomes were significantly different but in the direction opposite that for the typical alternative-outcomes effect. This finding provides a novel illustration of the importance that frequency-based regression effects might have for probability judgments. It is not necessarily the case that the regression effects and the alternative-outcomes effect will always operate in opposite directions, as they did here. However, although the heuristic comparison account of the alternative-outcomes effect suggests that the frequency of the strongest alternative has disproportionate weight in focal-likelihood judgments, regression effects will always reduce the perceived frequency of that strongest alternative more than any other alternative. This observation suggests that in tasks represented by this learning paradigm (and in other tasks in which information loss is an issue), regression effects and heuristic comparison processes will tend to have conflicting consequences.

REFERENCES

- ATTNEAVE, F. (1953). Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology*, *46*, 81-86.
- BETSCH, T., SIEBLER, F., MARZ, P., HORMUTH, S., & DICKENBERGER, D. (1999). The moderating role of category salience and category focus in judgments of set size and frequency of occurrence. *Personality & Social Psychology Bulletin*, *25*, 463-481.
- BROWN, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 1539-1553.
- BROWN, N. R. (1997). Context memory and the selection of frequency estimation strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *23*, 898-914.
- FIEDLER, K., & ARMBRUSTER, T. (1994). Two halves may be more than one whole: Category-split effects on frequency illusions. *Journal of Personality & Social Psychology*, *66*, 633-645.
- GONZALEZ, M., & FRENCK-MESTRE, C. (1993). Determinants of numerical versus verbal probabilities. *Acta Psychologica*, *83*, 33-51.
- GREENE, R. L. (1986). Effects of intentionality and strategy on memory for frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *12*, 489-495.
- GREENE, R. L. (1989). On the relationship between categorical frequency estimation and cued recall. *Memory & Cognition*, *17*, 235-239.
- HASHER, L., & ZACKS, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, *39*, 1372-1388.
- HINTZMAN, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, *95*, 528-551.
- HINTZMAN, D. L. (2001). Similarity, global matching, and judgments of frequency. *Memory & Cognition*, *29*, 547-556.
- HOWELL, W. C. (1973). Representation of frequency in memory. *Psychological Bulletin*, *80*, 44-53.
- JONIDES, J., & NAVEH-BENJAMIN, M. (1987). Estimating frequency of occurrence. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *13*, 230-240.
- MAKI, R. H., & OSTBY, R. S. (1987). Effects of level of processing and

- rehearsal on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **13**, 151-163.
- MANIS, M., SHEDLER, J., JONIDES, J., & NELSON, T. E. (1993). Availability heuristic in judgments of set size and frequency of occurrence. *Journal of Personality & Social Psychology*, **65**, 448-457.
- SANBONMATSU, D. M., POSAVAC, S. S., & STASNEY, R. (1997). The subjective beliefs underlying probability overestimation. *Journal of Experimental Social Psychology*, **33**, 276-295.
- SEDLMEIER, P., HERTWIG, R., & GIGERENZER, G. (1998). Are judgments of the positional frequencies of letters systematically biased due to availability? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **24**, 754-770.
- 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. (2001). When equal chances = good chances: Verbal probabilities and the equiprobability effect. *Organizational Behavior & Human Decision Processes*, **85**, 77-108.
- TVERSKY, A., & KOEHLER, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, **101**, 547-567.
- VAN WALLENDIAEL, L. R., & HASTIE, R. (1990). Tracing the footsteps of Sherlock Holmes: Cognitive representations of hypothesis testing. *Memory & 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 & Human Decision Processes*, **81**, 195-225.
- 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 & Social Psychology*, **75**, 1411-1423.
- WINDSCHITL, P. D., & YOUNG, M. E. (2001). The influence of alternative outcomes on gut-level perceptions of certainty. *Organizational Behavior & Human Decision Processes*, **85**, 109-134.
- WRIGHT, G., & WHALLEY, 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: Reidel.

(Manuscript received July 25, 2001;
revision accepted for publication January 9, 2002.)