Journal of Experimental Psychology:
Learning, Memory, and Cognition
2004, Vol. 30, No. 1, 198–215

Copyright 2004 by the American Psychological Association, Inc.
0278-7393/04/$12.00 DOI: 10.1037/0278-7393.30.1.198

The Dud-Alternative Effect in Likelihood Judgment

Paul D. Windschitl and John R. Chambers
University of Iowa

The judged likelihood of a focal outcome should generally decrease as the list of alternative possibilities increases. For example, the likelihood that a runner will win a race goes down when 2 new entries are added to the field. However, 6 experiments demonstrate that the presence of implausible alternatives (*duds*) often increases the judged likelihood of a focal outcome. This *dud-alternative effect* was detected for judgments involving uncertainty about trivia facts and stochastic events. Nonnumeric likelihood measures and betting measures reliably detected the effect, but numeric likelihood measures did not. Time pressure increased the magnitude of the effect. The results were consistent with a contrast-effect account: The inclusion of duds increases the perceived strength of the evidence for the focal outcome, thereby affecting its judged likelihood.

People must often assess the likelihood that a particular outcome, rather than one of multiple alternative outcomes, will occur. For example, a newspaper reader might assess the likelihood that the Democratic nominee for the 2004 U.S. presidential election will be John Kerry rather than Joe Lieberman, John Edwards, Richard Gephardt, or some other candidate. An admissions counselor might estimate the chances that a student will major in biology rather than psychology or chemistry. An eyewitness might estimate the chance that Suspect X was the culprit rather than Suspects Y or Z. This article concerns how people make such judgments. More specifically, this article concerns how people's evaluations of evidence for multiple alternative outcomes ultimately shape their likelihood judgments about a focal outcome.

One account for these judgments assumes that people execute a set of processing steps related to normative probability theory. Specifically, people might assess the support (i.e., the perceived strength of the evidence) for the focal outcome and for each of the alternatives, then aggregate the support for the alternatives, and finally compute a subjective probability by comparing the support for the focal outcome to the aggregated support for the alternatives. A descriptive theory of subjective probability—support theory—bears some resemblance to this type of account (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994). Support theory suggests that the judged probability, $P(A, \bar{A})$, 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, $s(\bar{A})$:

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

Paul D. Windschitl and John R. Chambers, Department of Psychology, University of Iowa.

This work was supported by Grant SES 99-11245 to Paul D. Windschitl from the National Science Foundation. We thank Michael E. Young and Michael R. P. Dougherty for their comments on a draft of this article, and we thank Zlatan Krizan for his work on the computer program for Experiment 6.

Correspondence concerning this article should be addressed to Paul D. Windschitl, Department of Psychology, University of Iowa, Iowa City, IA 52242. E-mail: paul-windschitl@uiowa.edu

A key assertion of the theory is that the perceived support for the residual, $s(\bar{A})$, is often less than the sum of the perceived support for the individual hypotheses that make up the residual. For example, people's support judgments for the hypothesis that one of John Kerry's competitors will win the 2004 Democratic nomination would be smaller than the sum of their separate support judgments for Joe Lieberman, John Edwards, Richard Gephardt, and all of the others. However, important for the present work is the fact that support theory, as it was originally proposed, does not further specify how support assessments for individual alternative hypotheses are treated or integrated when a person is gauging the overall support for the residual.

Recent research investigating how support for individual alternative hypotheses influences likelihood judgments has produced findings that are not anticipated by a normative account or by support theory (Teigen, 2001; Windschitl & Wells, 1998; Windschitl & Young, 2001; Windschitl, Young, & Jenson, 2002). Specifically, demonstrations of *alternative-outcomes effects* have revealed that people's perceptions of certainty regarding a focal outcome will vary as a function of how evidence for the alternative outcomes is distributed, even when the overall amount of evidence for the alternative outcomes is held constant. For example, in one study, participants' intuitive certainty about winning was greater for a raffle in which they held 17 tickets and others held 8, 8, 9, 8, and 9 tickets than it was for a raffle in which they held 17 tickets and others held 7, 6, 16, 6, and 7 tickets (Windschitl & Young, 2001).

Demonstrations of alternative-outcomes effects gave rise to a new type of account for how support for individual alternative hypotheses influences likelihood judgments. Windschitl and Wells (1998) proposed that a key determinant of likelihood judgments in multialternative cases is a heuristic comparison process in which the evidence for the focal outcome is compared with the evidence for each of the individual alternative hypotheses. The key component of this account that explains alternative-outcomes effects is that the comparison between the focal outcome and the strongest alternative outcome plays a disproportionate role (relative to other comparisons) in determining the perceived certainty of the focal outcome. The more this comparison favors the focal outcome, or the less it favors the strongest alternative outcome, the greater the

perceived likelihood of the focal outcome. The processes driving the alternative-outcomes effect are assumed to be relatively efficient in that they do not require an effortful aggregation of support across alternatives. Nevertheless, the comparison between the evidence for the focal and strongest alternative outcome can serve as a roughly accurate guide as to whether one should feel optimistic or pessimistic about the possibility of the focal outcome (see Windschitl & Young, 2001).[1]

In the present article, we identify and investigate another effect relevant to likelihood judgments in multialternative cases. Like the alternative-outcomes effect, this new effect is counternormative and is not anticipated by support theory. We refer to the effect as the *dud-alternative effect* because it shows that adding very weak alternatives—*duds*—to a likelihood question can increase the judged likelihood of the focal outcome.

Imagine, for example, that a person reads "Either Calcutta, India; Cincinnati, Ohio; Nairobi, Kenya; or Moscow, Russia is below the equator. How likely is it Nairobi?" The person may immediately recognize that Cincinnati and Moscow are so implausible that they can be ignored. Nevertheless, those duds may still have some influence on the judged likelihood that Nairobi is the city that is below the equator. A normative account might predict that the duds could garner some minimal support and would therefore ultimately reduce the judged likelihood of the focal hypothesis. Our prediction, however, was that the inclusion of two duds, compared with their exclusion from the question, would tend to increase the judged likelihood of the focal hypothesis.[2]

We present two related but distinct accounts for expecting a dud-alternative effect. Our first account is the *contrast account.* This account, like the account offered for the alternative-outcomes effect, begins with an assumption that people compare the evidence for the focal outcome to evidence for the individual alternative outcomes. Although the alternative-outcomes effect illustrates that the comparison between the focal and the strongest alternative outcome has a disproportionately strong role in determining the perceived certainty of the focal outcome, the comparisons between the focal and weaker outcomes are not irrelevant. These alternative outcomes, even the weak ones, form a local context against which the evidence for the focal outcome is compared. The inclusion of duds in this local context increases the number of times that the focal outcome compares very favorably to individual alternatives, thereby increasing the perceived strength and ultimately the judged likelihood of the focal outcome. For example, the inclusion of Cincinnati and Moscow in the equator question increases the number of times (by 2) that Nairobi compares very favorably against individual alternatives. By contrast to these duds, the evidence for Nairobi seems strong, which ultimately increases the judged likelihood that Nairobi is the city that is below the equator.[3]

Our contrast account bears some relation to value-shift explanations for the effects of decoys in choice sets (see, e.g., Ariely & Wallsten, 1995; Huber, Payne, & Puto, 1982; Simonson & Tversky, 1992; Wedell, 1991; Wedell & Pettibone, 1996). According to those explanations, decoys can increase people's tendency to pick a dominating option because the subjective value of an attribute from that option is enhanced by the presence of a decoy (see, e.g., Wedell & Pettibone, 1996). Our contrast account of the dud-alternative effect applies a somewhat similar logic to a likelihood judgment task and suggests that the enhanced subjective evaluation of evidence for a focal option is what mediates the impact of duds on a focal-option likelihood judgment.

A second account for expecting a dud-alternative effect is referred to as the *averaged-residual account.* This account, like support theory, assumes that participants do not make pairwise comparisons between the focal and the individual alternative outcomes. Instead, when participants judge focal-option likelihood, they first judge the support offered by the evidence for the focal option and the support offered by some aggregated representation of the evidence for the alternatives comprising the residual. They then compare the support for the focal option to support for the residual. This averaged-residual account is closely related to support theory, but the critical feature of this account is that the aggregated representation for the residual is based on an averaging of the evidence for the individual alternatives. Because of this averaging, the judged support for the residual is lower when duds are present rather than absent, thereby increasing the judged likelihood of the focal option.

If participants do attempt to aggregate evidence for a multicomponent residual, is it plausible to expect that they average the evidence for the individual components? Research on the disjunction fallacy provides reasons to suspect they might. Carlson and Yates (1989) found results consistent with the idea that participants sometimes average component likelihoods when judging disjunction likelihoods (for related work see Bar-Hillel & Neter, 1993; Rottenstreich, Brenner, & Sood, 1999). In that research, the disjunction was in a focal position (i.e., it was the hypothesis that was explicitly asked about). However, one might expect that a similar averaging process would occur if the disjunction served as the residual, as is the case for our dud-present questions (e.g., the "Calcutta, Cincinnati, or Moscow" residual in our above-mentioned question).

Goals of the Research

In this article, we describe six experiments. In the general methods and procedures for these experiments, we had participants respond to likelihood questions that always contained a focal

[1] The heuristic-comparison account is not intended as an exclusive account of how people make likelihood judgments. It is assumed that people can and often do engage in likelihood-judgment processes that resemble a normative account; however, under many circumstances, their judgments are driven primarily by the results of relatively effortless pairwise comparisons between the focal and alternative outcomes.

[2] Although we use terms like *increase* when discussing the dud-alternative effect (e.g., "the inclusion of duds can cause an increase in judged likelihood"), our work does not focus on belief updating (see, e.g., Hogarth & Einhorn, 1992; Robinson & Hastie, 1985). In our experiments, participants responded to either a dud-absent or dud-present version of a question; they did not respond to a dud-absent version and then update when duds were added.

[3] Range-frequency theory could be used as a more detailed explanation for why including or excluding duds could influence the judged strength of a focal option, resulting in a contrast effect (see Parducci, 1965). In short, the theory specifies that a judgment of a focal stimulus is a function of where the relevant value of the focal stimulus falls within the range of values for the context stimuli (range principle) and what proportion of context stimuli have values that fall below that of the focal stimulus (frequency principle).

hypothesis and a nondud alternative; sometimes those hypotheses were accompanied by a small number of dud alternatives (the dud-present condition) and sometimes they were not (the dud-absent condition). Our first goal was to determine whether a dud-alternative effect could be demonstrated in a reliable fashion. To determine whether the effect has broad relevance to likelihood judgment processes rather than narrow relevance to one specific judgment task, the effect was tested in two vastly different types of likelihood-judgment paradigms. We also investigated the differential sensitivity of two types of likelihood scales to the dud-alternative effect, and we tested whether the effect would influence the amount of money that participants reported they would wager in a hypothetical betting task. Finally, we tested whether time pressure would enhance the magnitude of the effect.

In this article, we do not attempt to establish either the contrast account or the average-residual account as an exclusive explanation for the dud-alternative effect. In fact, we suspect that the mechanisms described in both accounts may play a role in producing dud-alternative effects and that their relative contributions likely vary depending on a variety of task factors. Nevertheless, we conducted this research with the contrast account in mind, and a main goal of one of our experiments was to establish evidence that was uniquely consistent with that account.

Experiment 1

Experiment 1 provided an initial test of the dud-alternative effect. Participants responded to numerous questions, each of which asked about the likelihood that a specified focal option was the most frequently selected option in a recent survey. Some of the questions that participants encountered were in a dud-present state (i.e., dud options were included in the option list) and some were in a dud-absent state (i.e., dud options were not included). We expected that the judged likelihood of a focal option would be higher when a question was seen in its dud-present rather than dud-absent state. After providing likelihood estimates for the focal options, participants encountered the questions again in a second questionnaire and provided likelihood estimates for all options in each question. This second questionnaire was used to test whether all of the nondud options—not just the focal options from the first questionnaire—would have higher judged likelihoods when duds were present versus when they were absent.

From where did we get our duds? We began with the important assumption that for an added dud to produce a detectable dud-alternative effect, the evidence for the dud would need to be extremely weak relative to the evidence for the focal outcome. We made this assumption for two interrelated reasons. First, research on contrast effects has revealed that, as a general rule, the magnitude of a contrast effect produced by a contextual stimulus will increase as a function of the extent to which that contextual stimulus is extreme and distinct from the target stimulus on the judgment dimension (see, e.g., Helson, 1964; Herr, 1986). Hence, adding Havana, Cuba to the option list in our earlier example involving Nairobi might not produce a detectable dud-alternative effect because the latitudinal positions of Havana and Nairobi do not differ enough. Second, and perhaps more obviously, to the extent that a set of respondents considers an added alternative plausible or nonnegligible, the judged likelihood of the focal outcome would be reduced, not increased. In an attempt to ensure that our duds would be perceived as very weak, we conducted pilot tests in which undergraduate laboratory assistants rated the extent to which possible options on likelihood questions would be perceived as highly implausible to most undergraduate research participants. The data from these pilot tests then guided our selection of the duds and nonduds to use in our questions.

Method

Participants. The participants were 44 students from Elementary Psychology classes at the University of Iowa who participated to fulfill a research exposure component of the course.

Procedure and materials. Participants began by completing either Form A or Form B of a questionnaire that contained 10 critical questions and 8 fillers (see Appendix A). In Form A, 5 of the critical questions were in their dud-present state, and 5 were in their dud-absent state; this assignment was reversed in Form B.

Each of the critical questions described a survey that, unbeknownst to participants, was fictional. The questions asked participants to judge the likelihood that a specified option was the most frequently selected one on the survey. The following is an example question in its dud-present state:

> In a recent survey, a random sample of U.S. children aged 7–10 was asked: "Which of the following is your favorite type of food for dinner?"
>
> Pizza
> Eggplant Parmesan
> Hamburger
> Grilled fish
>
> How likely do you think it is that *pizza* was the most frequently selected option?

All dud-present versions described two survey options that had been prejudged in pilot tests as at least somewhat plausible (e.g., "Pizza," "Hamburger") and two duds that had been prejudged in pilot tests as completely implausible (e.g., "Eggplant Parmesan," "Grilled fish"). One of the plausible (nondud) options was arbitrarily selected to serve as the focal option for the question. The dud-absent versions of the questions were identical to the dud-present versions, except that the two duds were not listed. Participants responded to all of the questions by circling 1 of 11 asterisks on a scale anchored by *extremely unlikely* (scored as 1) and *extremely likely* (scored as 11).

After completing the first questionnaire, which contained our main dependent measures, participants completed a second questionnaire that asked them to give likelihood responses to all of the possible options from the same questions to which they had responded in the first questionnaire. Below is an example of how the dud-present version of the favorite-food question appeared in the second questionnaire.

> Earlier you read that a random sample of U.S. children aged 7–10 was asked: "Which of the following is your favorite type of food for dinner?" (Pizza—Hamburger—Eggplant Parmesan—Grilled fish). For each response option, indicate what you think is the chance that it was the most frequently chosen response.

We used a different response format for the second questionnaire than we had used for the first: Participants responded by circling a number from 0 (*no chance*) to 6 (*high chance*). This format removed the possibility that participants would simply recall and reuse the same answer they had given on the first questionnaire when responding to focal options on the second questionnaire.

Results and Discussion

First questionnaire data. For each participant, we calculated his or her average response to the five questions seen in a dud-

present version and his or her average response to the five questions seen in a dud-absent version. These values were submitted to a 2 (dud present or absent) × 2 (Form A or B) mixed-model analysis of variance (ANOVA) with the first factor as a within-participants factor. The means for these values are shown in Table 1. The key finding was a significant main effect for the presence or absence of duds, $F(1, 42) = 33.37$, $p < .01$. The direction of the effect was as predicted; specifically, the estimated likelihoods of the focal options were significantly higher for the dud-present versions ($M = 9.37$, $SD = 0.81$) than they were for the dud-absent versions ($M = 8.55$, $SD = 0.79$). The form main effect was nonsignificant, $F(1, 42) = 2.45$, $p = .13$, and although the interaction term was significant, $F(1, 42) = 6.60$, $p < .05$, this result is not particularly noteworthy because it simply reflects an arbitrary property of our counterbalancing scheme.[4]

The magnitude of the dud-alternative effect was substantial. In fact, the effect size ($d = 1.02$) would be classified as large according to Cohen's (1988) conventional values. Also, the differences between means for dud-present and dud-absent versions were in the predicted direction for 9 of the 10 critical questions, including the favorite-food question described in the *Method* section. (The mean judged likelihood for "Pizza" was 9.96, $SD = 1.15$, when "Eggplant Parmesan" and "Grilled fish" were included in the question but only 9.43, $SD = 1.47$, when they were omitted.) The consistency of the findings across items suggests that the effect does not reflect idiosyncratic features of a small set of items but instead reflects a general process by which the inclusion of dud alternatives increases the judged likelihood of a focal option.

Second questionnaire data. Did the presence of duds influence only the judged likelihood of the focal outcomes that we arbitrarily selected for the first questionnaire? In the second questionnaire, participants provided likelihood judgments for every option listed in a question, including the options that had served as the nondud, nonfocal options in the first questionnaire ("Hamburger" from our earlier example). We submitted the responses for these nondud options to the same type of 2 × 2 ANOVA described above and found a highly similar pattern of results. The form main effect was nonsignificant, $F(1, 42) = 1.54$, $p = .22$, whereas the interaction effect was significant, $F(1, 42) = 10.88$, $p < .01$. More important, the main effect for the presence or absence of duds was again significant, $F(1, 42) = 27.90$, $p < .01$; the estimated likelihoods for those options were significantly higher for the dud-present versions ($M = 4.70$, $SD = 0.64$) than they were for the dud-absent versions ($M = 4.25$, $SD = 0.71$). Hence, our selection of which options to treat as focal options did not play a critical role in the dud-alternative effect; the addition of duds to a question appears to influence responses about all of the nondud options in the question.

To confirm that the likelihood responses for the focal outcomes again showed a dud-alternative effect on the second questionnaire, we submitted those responses to a 2 × 2 ANOVA and found results similar to those for the first questionnaire. Most important, the main effect for the presence or absence of duds was again significant, $F(1, 42) = 31.79$, $p < .01$; likelihood estimates were higher for the dud-present versions ($M = 5.59$, $SD = 0.41$) than they were for the dud-absent versions ($M = 5.22$, $SD = 0.60$).

Finally, we examined the mean likelihood responses for the dud alternatives to evaluate whether our pilot testing was successful at identifying duds that most of our participants would find completely implausible. Optimally, our duds would have received responses at or near 0 (*no chance*). Although there were many such responses, our analyses suggest that we were not completely successful. One dud received a mean response of 3.05. The mean responses for the remaining 19 duds ranged from 0.86 to 2.54. The fact that some duds were not viewed as implausible by all participants likely worked against the predicted direction of our dud-alternative effects, making the effects that were detected all the more impressive.

Table 1
Mean Likelihood Judgments From Experiment 1 as a Function of Version and Form

	Version			
	Dud-absent		Dud-present	
Form	*M*	*SD*	*M*	*SD*
A	8.23	.75	9.40	.97
B	8.89	.70	9.33	.61
Overall	8.55	.79	9.37	.81

Experiment 2

According to our contrast account for the dud-alternative effect, the inclusion of duds increases the number of times that the evidence for the focal option compares quite favorably to that for individual alternative options. By way of a contrast effect, this enhances the perceived strength of the focal option and can thereby influence a person's certainty about that option. According to the averaged-residual account, the inclusion of duds increases the judged likelihood of the focal option because they lower the averaged support for the residual.

In Experiment 2, we tested an important alternative to both of these accounts. This alternative account, like our contrast account, assumes that the presence of duds can lead to contrast effects. However, this account assumes that the contrast effects merely reflect a change in the way in which participants used the response scales, not a change in their internal certainty regarding a focal option (for related discussions see, e.g., Campbell, Lewis, & Hunt, 1958; Krantz & Campbell, 1961; Manis, 1967; Manis & Armstrong, 1971; Simpson & Ostrom, 1976; Stevens, 1958; Upshaw, 1969). This may have happened in Experiment 1 if participants implicitly or explicitly treated the absolute likelihood questions (e.g., "How likely do you think it is that pizza was the most frequently selected option?") as relative likelihood questions (e.g., "Relative to the other listed options, how likely do you think it is that pizza was the most frequently selected option?"). Under the latter treatment, a scale label such as *extremely likely* could have had a different connotation when the only alternative option was "Hamburger" than when the alternatives included "Hamburger," "Eggplant Parmesan," and "Grilled fish."

[4] Specifically, the five questions we arbitrarily assigned to be dud-present in Form A and dud-absent in Form B received higher likelihood estimates on average than the five questions we assigned to be dud-absent in Form A and dud-present in Form B.

Although we suspect that participants in Experiment 1 did indeed engage in relative judgment processes when answering the likelihood questions, we also suspect that those relative judgment processes affected internal perceptions of certainty, not merely the interpretation or selection of responses on the likelihood scale. If so, we would expect that the presence or absence of duds would also affect responses that are mediated by perceptions of certainty, even if the scale on which those responses were made had externally meaningful reference points that would not be interpreted in relative terms. Hence, in Experiment 2, we replicated the methods of Experiment 1 except that half of the participants were asked to indicate a hypothetical bet (between $0 and $5) on the focal outcome of each question. We assumed that betting responses would be mediated by, and therefore reflect, participants' internal beliefs about the subjective likelihoods of the focal options. Because we assumed that duds influence internal perceptions of certainty, not merely the use of a subjective likelihood scale, we predicted that participants would bet more on the focal options when duds were present rather than absent.

The other half of the participants in Experiment 2 were asked to provide numeric subjective probability estimates (between 0% and 100%) for the focal options. Although we expected that the betting dependent measure would detect the same type of dud-alternative effect as did the measure used in Experiment 1, we did not expect the same for the numeric measure. This prediction follows from research showing that standard numeric measures are sometimes relatively insensitive to likelihood-judgment phenomena that can be detected with other types of measures, such as betting measures, choice measures, or measures involving verbal-likelihood response options (see, e.g., Kirkpatrick & Epstein, 1992; Teigen, 1988, 2001; Windschitl, 2000; Windschitl, Krizan, & Flugstad, 2003; Windschitl, Martin, & Flugstad, 2002; Windschitl & Wells, 1996, 1998). Research participants appear to be more concerned with accuracy and implementing formal rules when they are asked for standard numeric probability responses rather than nonnumeric likelihood responses (although see Windschitl, Young, & Jenson, 2002). Regarding the present experiment, we suspected that compared with participants in Experiment 1 and participants giving betting responses, participants giving numeric probability responses would engage in more deliberate and rule-based processing when generating their likelihood judgments and would be more cognizant of the constraints that complementarity and additivity place on likelihoods (see Windschitl, 2000). Those participants would be in a better position to appreciate that the presence of weak alternatives lowers rather than raises the likelihood of other hypotheses in a mutually exclusive and exhaustive set. This would mitigate the contrastive influence that the presence of duds can have on judgments of focal-outcome likelihood. Therefore, although we expected to find robust dud-alternative effects on betting measures, we did not expect these effects on numeric probability measures.

Method

Participants. The participants were 111 students from Elementary Psychology classes at the University of Iowa who participated to fulfill a research exposure component of the course.

Procedures and materials. Participants were randomly assigned to provide either bets or numeric probability estimates for the same set of questions used in the first questionnaire of Experiment 1. Participants in the betting condition were told to imagine that they could bet between $0 and $5 on each question in their questionnaire. The betting questions took the following form: "How much money ($0–$5) would you bet that X was the most frequently selected option?" Participants in the numeric-probability condition were told that they should provide a numeric chance estimate between 0% and 100% for each question. These questions took the following form: "What do you think is the chance that X was the most frequently selected option?" Initial instructions explained that a response of 25% would mean that there was a 1-in-4 chance that the specified option was the most frequently selected one on the survey.

Results

Although participants were randomly assigned to either the betting or numeric-probability conditions, we analyzed these conditions separately.

Bets. For participants in the betting condition, we calculated their average bet for the five questions that they saw in a dud-present version and their average bet for the five questions that they saw in a dud-absent version. As in Experiment 1, these values were submitted to a 2 (dud present or absent) × 2 (Form A or B) mixed-model ANOVA. The means for these values are shown in Table 2. The key finding was that the predicted main effect for the presence or absence of duds was significant, $F(1, 53) = 15.67$, $p < .01$; participants' bets on the focal options were significantly higher for the dud-present versions ($M = 3.77$, $SD = 0.84$) than they were for the dud-absent versions ($M = 3.36$, $SD = 1.06$). The form main effect was not significant, $F(1, 53) < 1$.

Numeric probability estimates. Estimates from the numeric-probability condition were submitted to the same type of mixed-model ANOVA as was used for the bets data. The key finding was that, as we predicted, the main effect for the presence or absence of duds was not significant, $F(1, 54) = 0.83$, $p = .37$; probability estimates for the focal outcomes were about the same for the dud-present versions ($M = 68.63$, $SD = 15.20$) as they were for the dud-absent versions ($M = 67.28$, $SD = 9.81$). The form main effect was not significant, $F(1, 54) = 2.46$, $p = .12$.

Discussion

As we expected, the presence versus absence of duds led to larger bets on the focal options but did not significantly affect the

Table 2
Mean Betting Estimates and Numeric Probability Judgments From Experiment 2 as a Function of Version and Form

	Version			
	Dud-absent		Dud-present	
Form	*M*	*SD*	*M*	*SD*
Betting estimates				
Form A	3.11	0.96	4.03	0.84
Form B	3.62	1.12	3.51	0.79
Overall	3.36	1.06	3.77	0.84
Numeric probability judgments				
Form A	68.02	7.89	72.60	13.34
Form B	66.54	11.52	64.65	16.12
Overall	67.28	9.81	68.63	15.20

judged numeric likelihoods of those options. The dud-alternative effect for bets is consistent with our contention that the presence of duds influences internal perceptions of certainty, not merely the selection of responses on the type of likelihood scale used in Experiment 1. In other words, adding a dud to a list of alternatives can actually enhance one's certainty about a focal option such that one would choose to bet more money on that option. (See Windschitl, Martin, & Flugstad, 2002, for related evidence that contrast effects detected on nonnumeric or subjective scales reflect changes to internal representations.)

The fact that duds did not significantly influence numeric probability estimates regarding focal options is consistent with a growing list of findings indicating that numeric measures are relatively insensitive to some likelihood phenomena that can be readily detected with verbal or nonnumeric measures (see, e.g., Kirkpatrick & Epstein, 1992; Teigen, 1988, 2001; Windschitl, 2000; Windschitl, Martin, & Flugstad, 2002; Windschitl & Wells, 1996, 1998). As suggested earlier, we suspect that in the case of the dud-alternative effect, soliciting numeric probability responses enhances people's awareness of complementarity and additivity constraints (Windschitl, 2000), thereby increasing an awareness that the presence of weak options—if having any effect at all—should lower rather than raise the likelihood of the focal option.

Experiment 3

In Experiments 1 and 2, the critical questions shared a notable property—each described another question that had supposedly appeared on a national survey (see Appendix A). Is the dud-alternative effect somehow dependent on this or other special properties of those questions? We suspected not, but nevertheless we thought it was important to replicate Experiments 1 and 2 using a new set of questions that did not involve descriptions of surveys. The old questions were also included in our replications. In these replications, we also made some minor modifications to the way in which focal options were identified in the questions.

Our replications were actually conducted as two separate experiments (a replication of Experiment 1 followed by a replication of Experiment 2), but for the sake of brevity, we describe them collectively as Experiment 3. Specifically, we refer to three conditions for Experiment 3: a nonnumeric-likelihood condition (the replication of Experiment 1) as well as a betting condition and a numeric-probability condition (the replication of Experiment 2). Given that the nonnumeric-likelihood condition was conducted separately from the other two, direct comparisons between conditions must be made with some caution.

Method

Participants. The participants were 163 students from Elementary Psychology classes at the University of Iowa (*n*s = 39, 63, and 61 in the nonnumeric-likelihood, betting, and numeric-probability conditions, respectively).

Procedure and materials. As in Experiments 1 and 2, participants began by completing either Form A or Form B of a questionnaire that contained 10 critical questions and 8 fillers. These questionnaires were the same as those used in Experiments 1 and 2, with only a minor modification that is described below. Next, participants completed either Form A or Form B of a second questionnaire that contained a new set of questions, 18 of which were critical ones and 3 of which were fillers. In Form A, 9 of the critical questions were in their dud-present state, and 9 were in their dud-absent state; this assignment was reversed in Form B.

Each of the critical questions, which are shown in Appendix B, began with a statement that one, and only one, of the listed options was the correct one. As was the case for the old questions, all dud-absent versions of new questions described two options that had been prejudged in pilot tests as at least somewhat plausible by a small group of research assistants. For each question, we had arbitrarily selected one of these plausible options to serve as the focal option. (Sometimes the focal option was the factually correct option, but sometimes it was not. This designation was not important, as our main hypotheses were not related to participant accuracy.) The dud-present versions included 2 additional options that, according to our pilot-tested assistants, would be viewed as completely implausible by most undergraduate participants.

Unlike Experiments 1 and 2, the focal option in a given question was always identified by a marker (>) that was printed beside the option. For example, the statement and options for one of the new questions (dud-present version) appeared as follows:

> One, and only one, of the following cities lies on the Mediterranean Sea.
>
> Cape Town, South Africa
> Cairo, Egypt
> > Naples, Italy
> Jakarta, Indonesia

Depending on the condition, participants were asked to provide either a nonnumeric-likelihood estimate (1 = *extremely unlikely*, 11 = *extremely likely*), a bet ($0–$5), or a numeric-probability estimate (0%–100%) that the marked option was the correct answer.

Results

Table 3 displays the relevant means for both the old and new sets of questions.

Old questions. We first examined whether the results for the old questions replicated the results from Experiments 1 and 2. We again calculated for each participant an average response to the five questions he or she saw in the dud-present version and an average response to the five questions he or she saw in a dud-absent version. These values were then analyzed using separate 2 (dud present or absent) × 2 (Form A or B) mixed-model ANOVAs (one for each of the three response-type conditions). To summarize those findings, the pattern of dud-alternative effects across the three conditions closely resembled the pattern detected across the analogous conditions of Experiments 1 and 2. Most important, the dud-alternative effect (i.e., the main effect for the presence or absence of duds) was significant in the nonnumeric-likelihood condition, $F(1, 37) = 16.27$, $p < .01$, and in the bets condition, $F(1, 61) = 6.18$, $p < .05$, but not quite significant in the numeric-probability condition, $F(1, 59) = 3.21$, $p = .08$.

New questions. The responses to the 18 new questions were analyzed in the same fashion. Again, a similar pattern was observed. The dud-alternative effect was significant in the nonnumeric-likelihood condition, $F(1, 37) = 11.97$, $p < .01$, and in the bets condition, $F(1, 61) = 17.21$, $p < .01$, but nonsignificant in the numeric-probability condition, $F(1, 59) = 0.84$, $p = .36$.

Table 3
Mean Nonnumeric Likelihood Judgments, Betting Estimates, and Numeric Probability Judgments From Experiment 3 as a Function of Version and Form

	Version			
	Dud-absent		Dud-present	
Response type	*M*	*SD*	*M*	*SD*
	Old questions			
Nonnumeric likelihood judgments				
Form A	8.42	1.57	9.45	1.37
Form B	8.54	0.90	8.69	1.23
Overall	8.48	1.26	9.06	1.34
Betting estimates				
Form A	3.48	0.67	4.01	0.60
Form B	4.06	0.44	3.86	0.46
Overall	3.78	0.63	3.93	0.53
Numeric probability judgments				
Form A	68.34	10.97	72.72	17.06
Form B	66.54	9.85	67.91	16.54
Overall	67.45	10.39	70.35	16.84
	New questions			
Nonnumeric likelihood judgments				
Form A	6.64	1.12	8.54	1.26
Form B	7.94	1.03	7.63	0.89
Overall	7.31	1.25	8.07	1.16
Betting estimates				
Form A	2.55	0.63	3.28	0.61
Form B	3.28	0.63	3.29	0.48
Overall	2.92	0.72	3.28	0.54
Numeric probability judgments				
Form A	54.99	10.07	62.24	13.90
Form B	60.41	8.10	56.10	14.57
Overall	57.66	9.48	59.22	14.45

Discussion

The results of Experiment 3 generalize the dud-alternative effect beyond the question set that was used in Experiments 1 and 2. The magnitude of the dud-alternative effects detected on nonnumeric likelihood measures in this experiment ($d = .45$ for old questions and $d = .63$ for new questions) was somewhat smaller than that detected in Experiment 1 ($d = 1.02$), but the effects are nevertheless notable in size, especially given the nonnormative nature of the effect. Perhaps more important, the results of this experiment confirm the pattern of findings across the different dependent variables that were used in Experiments 1 and 2. Numeric-probability measures—as opposed to nonnumeric measures and betting measures—are relatively insensitive to dud-alternative effects.

Experiment 4

The evidence presented thus far has been generally consistent with both the contrast account and the averaged-residual account. Although we assume that an averaging process, like that described in the averaged-residual account, might play some role in producing dud-alternative effects, we view the contrast account as an equally tenable, if not more tenable, explanation for the dud-alternative effect observed thus far. In Experiment 4, we sought to provide more direct and unique evidence consistent with the contrast account.

A key distinction between the two accounts is that the averaged-residual account does not predict that the presence of duds would affect the perceived strength of (i.e., support for) the focal option. Instead, that account simply assumes that the presence of duds lowers the perceived support for the residual, thereby increasing the judged likelihood of the focal option. According to the contrast account, the list of options in a likelihood question essentially serves as a local context against which the evidence for the focal outcome can be compared. Although people could compare the evidence for the focal option to evidence for unlisted self-generated options, we suspect that comparisons with the salient options that are explicitly mentioned play a dominant role. Hence, when two duds are added to a local context, this increases the number of salient pairwise comparisons in which the focal option compares quite favorably to alternatives, which thereby enhances the perceived strength and likelihood of the focal option.

If the dud-alternative effect is at least partially due to the processes described by the contrast account, then we should expect to find evidence that explicitly mentioning duds within a local context can enhance strength assessments of a focal option, even when the dependent measures have nothing to do with likelihood judgment. In Experiment 4, participants read about the same

survey questions and options that were used in Experiments 1 and 2.[5] However, instead of being asked to make a likelihood judgment for each question, they were asked to rate the strength of the focal option relative to all possible options of its type, not just those listed on the survey. The precise nature of this strength dimension necessarily depended on the survey question. For example, for the question about the most preferred food among children, the strength rating for a food was, more precisely, a rating of how much the typical child likes that food. We expected that, in general, the strength ratings of the focal options would be higher for the dud-present versus the dud-absent versions of the critical questions.

Method

Participants. The participants were 38 students from Elementary Psychology classes at the University of Iowa who participated to fulfill a research exposure component of the course.

Procedures and materials. Initial instructions informed participants that they would be reading about survey items and that for each, they should do two things:

> First, take a close look at each of the possible response options that were included on the survey; you will be asked to recall from memory many of these response options later. Second, after you have closely examined the survey options, read and respond to the question.

Participants were told about a memory test for the sole purpose of ensuring that they carefully read all of the response options (i.e., the local context) before answering the question.

Participants then read about the same survey questions and options that were used in Experiments 1 and 2, but they provided strength ratings rather than likelihood judgments for the focal outcomes. The dimension on which strength was rated depended on the nature of the survey item. The kids-food question read as follows for the dud-present version:

> In a recent survey, a random sample of U.S. children aged 7–10 was asked: 'Which of the following is your favorite type of food for dinner?—pizza, eggplant parmesan, hamburger, grilled fish.' Compared to all dinner foods, how much does the typical U.S. child aged 7–10 like *pizza?* (1 = *not at all,* 9 = *very much*).

As other examples, the question regarding music groups read: " 'Compared to all modern rock groups, how enjoyable is the Dave Matthews Band to the typical U.S. female college student?' (1 = *not enjoyable,* 9 = *very enjoyable*)." The question regarding shoe brands read: " 'Compared to all shoe brands on the U.S. market, how attractive is Nike to the typical U.S. teenage boy?' (1 = *not attractive,* 9 = *very attractive*)."

Results and Discussion

Similar to the analyses for Experiments 1–3, we calculated each participant's average strength rating for the five questions that he or she saw in a dud-present version and the average rating for the five questions that he or she saw in a dud-absent version. These values were submitted to a 2 (dud present or absent) × 2 (Form A or B) mixed-model ANOVA. The means for these values are shown in Table 4. Consistent with our hypothesis, participants gave higher strength ratings for focal outcomes in a dud-present version ($M = 7.62$, $SD = 0.97$) than they did for focal outcomes in a dud-absent version ($M = 7.20$, $SD = 0.91$), $F(1, 38) = 7.33$, $p = .01$. Neither the form main effect nor the interaction was significant (both Fs < 1).

Table 4
Mean Strength Judgments from Experiment 4 as a Function of Version and Form

	Version			
	Dud-absent		Dud-present	
Form	*M*	*SD*	*M*	*SD*
Form A	7.15	0.73	7.56	0.72
Form B	7.26	1.09	7.68	1.18
Overall	7.20	0.92	7.62	0.97

The dependent measures in this experiment asked participants to compare the strength of a focal item (e.g., pizza) to all possible items of its type (which should include not only eggplant parmesan, hamburger, and grilled fish but also macaroni and cheese, spaghetti, liver and onions, etc.). In other words, participants should have given equal weight to all possible comparison standards (all dinner foods), regardless of whether the question was a dud-absent or dud-present version. However, the results make it clear that the presence of duds influenced the actual comparison standards that participants used to evaluate the strength of the focal outcome. Eggplant parmesan and grilled fish, for example, were presumably more influential as comparison standards in the dud-present condition than they were in the dud-absent condition. A related possibility is that the presence of these duds extended the range of foods (in the "yucky foods" direction) that were considered as comparisons when participants evaluated pizza. Regardless of how the influence of duds in this experiment is best characterized, we suspect that duds can have a similar influence when participants are making internal assessments of strength en route to judging likelihood. For example, when thinking about whether pizza is likely to be the most-liked food among kids, pizza seems even more likable when the likelihood question also mentions eggplant parmesan and grilled fish.

Experiment 5

In Experiment 5, we investigated the dud-alternative effect using a completely different paradigm from that used in Experiments 1–4. Testing the effect in a different paradigm is important because we assumed that the processes driving the dud-alternative effect (at least those described by the contrast account) are general processes that apply to almost any form of likelihood judgment involving multiple alternative hypotheses, not just judgments about trivia questions. For example, we assumed that these processes apply to not only judgments involving uncertainty that is based on lack of personal knowledge about a fact (such as when a person is not sure about which answer is the correct answer to a trivia question) but also judgments involving uncertainty that is based on an inability to foresee how a stochastic event will play out (such as who the winner of a raffle will be). In Experiment 5,

[5] Although we describe Experiment 4 after Experiment 3, Experiment 4 was actually conducted prior to Experiment 3 and, hence, prior to the creation of the new items used in Experiment 3. This is why strength judgments for the new items were not tested in Experiment 4.

we used a paradigm that tested the latter type of likelihood judgment.

Participants viewed numerous representations of raffles, each of which indicated how many tickets they and other players held in a hypothetical raffle (see Figure 1). For each raffle, participants indicated how good they would feel at a "gut level" about their chances of winning. This paradigm, including the gut-level response scale, has been used successfully to investigate alternative-outcomes effects (Windschitl & Young, 2001). As in studies of the alternative-outcomes effect, we assumed that it would be important to encourage people to provide their gut-level responses (and record them on nonnumeric scales); otherwise participants would be likely to engage a rule-based calculation for generating their responses, thus precluding the detection of dud-alternative effects (see Windschitl & Wells, 1998, but see also Windschitl, Young, & Jenson, 2002). The key manipulation in the experiment was whether the nondud players in a raffle were accompanied by no dud players (the baseline version), some dud players (a dud-present version), or some "mediocre" players (a mediocre-present version). Figure 1 displays an example of each of these three versions for a raffle used in the experiment.

The contrast account provides predictions for this paradigm that are relatively straightforward. We expected that participants would report greater optimism about winning the dud-present raffles than they would about winning the baseline raffles, even though they should be more optimistic about winning the latter raffles than they should be about winning the former. According to the contrast account, comparisons between evidence for the focal outcome (i.e., the participant winning) and individual alternatives would be generally quite favorable in the dud-present condition but not the baseline condition, thus enhancing—via a contrast effect—the perceived strength and likelihood of the focal outcome. At the same time, we expected that the inclusion of mediocre alternatives would not necessarily lead to a significant increase in the perceived likelihood of winning. Although the mediocre alternatives might slightly boost—via a contrast effect—the perceived strength of the focal outcome, this boost would be smaller than the one in the dud-present condition and offset by the fact that the additional mediocre alternatives would not be viewed as negligible (see our discussion in Experiment 1).

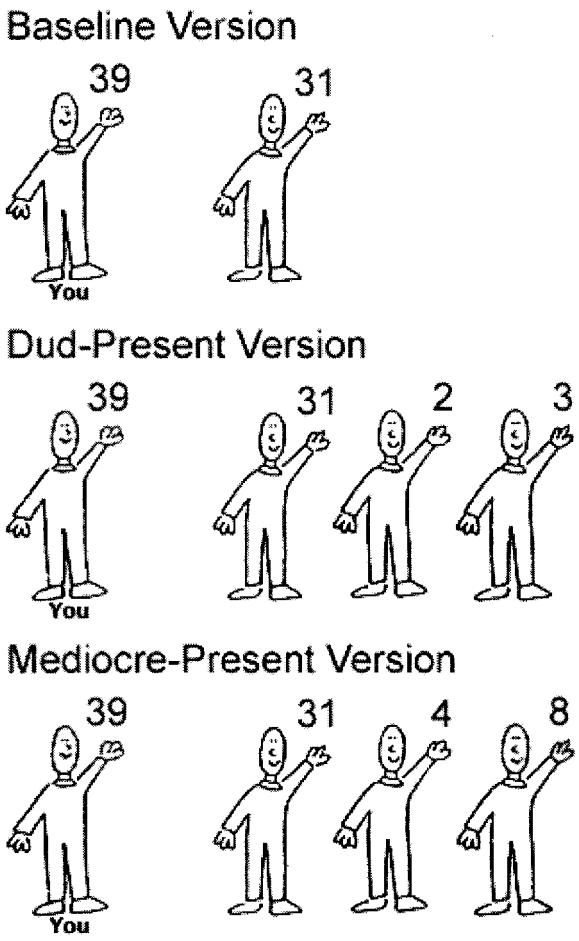

Figure 1. Three raffles from one of the sets (no. 4) used in Experiment 5.

The averaged-residual account makes a slightly different set of predictions. Like the contrast account, it predicts that participants would be more optimistic about winning the dud-present raffles than they would about winning the baseline raffles, because the presence of duds would reduce the averaged strength of the residual. However, the account would also predict that participants would be more optimistic about winning the mediocre-present raffles than they would about winning the baseline raffles. Adding two mediocre players would substantially reduce the averaged strength of the residual. For example, inserting mediocre players holding 4 and 8 tickets into a residual containing a player holding 31 tickets (see Figure 1) substantially reduces the averaged overall strength of that residual (from 31 to 14.3 tickets), thereby increasing a participant's optimism about winning.

Method

Participants. The participants were 41 students from Elementary Psychology classes at the University of Iowa who participated to fulfill a research exposure component of the course.

Procedure. Each participant received a booklet that contained representations of 24 critical raffles and 9 filler raffles. The order of the 24 raffles was randomized (separately for each participant) with a constraint that ensured that raffles from a given set were separated by several pages in the booklet (see below for information on sets). For each raffle, participants responded to the following: "At a gut level, how would you feel about your chances of winning this raffle?" Participants responded by circling one of nine asterisks on a scale anchored by *not good at all* (scored as 1) and *very good* (scored as 9). 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."

The raffles. Table 5 displays the 24 critical raffles, which can be thought of conceptually as eight sets of 3 raffles.[6] In each set, there was a baseline, dud-present, and mediocre-present raffle. The baseline raffles from each set, which contained only the participant and one other player,

[6] The filler raffles, which are not included in Table 5, contained exactly three players each. These fillers were included to reduce the chances that participants, while completing their booklets, would find it peculiar that each raffle contained either two or four players (and sometimes five; see Table 5). The fillers also helped to decrease the chances that participants would notice similarities between raffles in a given set.

Table 5
Ticket Distributions, Objective Probability of Winning, and Mean Optimism in Experiment 5 as a Function of Raffle Set and Type

Raffle set and type	Number of tickets held					Objective probability	Optimism	
	"You"	Other players					*M*	*SD*
Set 1								
Baseline	14	16				0.467	5.19	1.70
Dud present	14	2	16	1		0.424	5.21	1.77
Mediocre present	14	6	16	5		0.341	5.17	1.50
Set 2								
Baseline	23	15				0.605	6.67	1.39
Dud present	23	15	3	2		0.535	6.70	1.55
Mediocre present	23	15	6	6		0.460	6.64	1.43
Set 3								
Baseline	12	14				0.462	5.32	1.79
Dud present	12	1	2	14		0.414	5.79	1.54
Mediocre present	12	3	4	14		0.364	5.49	1.76
Set 4								
Baseline	39	31				0.557	6.23	1.41
Dud present	39	31	2	3		0.520	6.81	1.57
Mediocre present	39	31	4	8		0.476	6.45	1.56
Set 5								
Baseline	17	7				0.708	7.73	0.86
Dud present	17	1	7	1	2	0.607	8.07	1.02
Mediocre present	17	3	7	3	4	0.500	7.11	1.52
Set 6								
Baseline	20	14				0.588	6.67	1.39
Dud present	20	2	14	1		0.541	6.86	1.41
Mediocre present	20	6	14	6		0.435	6.17	1.50
Set 7								
Baseline	45	53				0.459	4.65	1.91
Dud present	45	3	4	53	3	0.417	5.79	1.73
Mediocre present	45	23	28	53	20	0.266	4.79	2.03
Set 8								
Baseline	41	36				0.532	6.24	1.34
Dud present	41	36	2	2	1	0.500	6.48	1.70
Mediocre present	41	36	29	22	22	0.273	5.36	2.05
Overall								
Baseline							6.09	1.07
Dud present							6.46	1.13
Mediocre present							5.90	1.22

were constructed somewhat arbitrarily but with the constraint that the number of tickets held by the participant and the other player were roughly comparable. A dud-present version was constructed by adding two or three players who had very few tickets relative to the focal player (i.e., the participant). A mediocre-present version was constructed by adding two or three players who had few tickets but more than those held by the dud players in the dud-present condition.

For each raffle, the participant's caricature—labeled "you"—and his or her number of tickets always appeared on the far left, as in Figure 1. The locations of the other types of players varied, as shown in Table 5. Previous research using this paradigm to test alternative-outcomes effects has revealed that the locations of strong versus weak alternatives in the raffle representations have little if any impact on optimism judgments (see Windschitl & Young, 2001).

Results and Discussion

Table 5 also displays the mean responses for the 24 raffles. The responses were submitted to a repeated-measures ANOVA with raffle type (baseline, dud-present, or mediocre-present) and raffle set (1–8) as within-participant factors. Not surprisingly, the main effect for raffle set was significant, $F(7, 34) = 31.00$, $p < .01$. The Raffle Set × Raffle Type interaction was also significant, $F(14, 27) = 3.22$, $p < .01$. Most important, however, the main effect for raffle type was significant, $F(2, 39) = 18.17$, $p < .01$.

A planned comparison revealed that, as we expected, participants were significantly more optimistic about winning the dud-present raffles ($M = 6.46$, $SD = 1.13$) than they were about winning the baseline raffles ($M = 6.09$, $SD = 1.07$), $F(1, 40) = 9.21$, $p < .01$, $d = .34$. Hence, even though the addition of duds slightly decreased the objective probability of winning, it significantly increased subjective optimism about winning. The same ordering of means for the responses to these two raffle types was observed for all eight sets, attesting to the consistency of the dud-alternative effect. Another planned comparison revealed that participants were not significantly more optimistic in the mediocre-present condition ($M = 5.90$, $SD = 1.22$) than they were in the baseline condition ($M = 6.09$, $SD = 1.07$), $F(1, 40) < 1$. This finding is consistent with our contention that duds need to be quite weak to produce detectable dud-alternative effects.

These findings extend the applicability of the dud-alternative effect to an entirely new type of likelihood judgment. Not only can the addition of duds increase the judged likelihood of an option in a task involving uncertainty about a fact or a completed event (as in Experiments 1–3) but the addition of duds can also have a similar effect in a task involving uncertainty about the future outcome of a stochastic event, such as a raffle. This bolsters our claim that the processes underlying the effect have broad applicability to various types of likelihood judgments.

The observed pattern of results is consistent with the contrast account but not with the averaged-residual account. Had participants been using a pure averaging strategy when judging the evidence for the residual, they would have expressed more optimism about winning the mediocre-present raffles than they did about winning the baseline raffles. As we discussed earlier, including the mediocre players necessarily caused a substantial reduction in the average number of tickets held by players in the residual, which should have, but did not, boost participants' optimism.

However, this pure version of the averaged-residual account could be slightly modified to account for the fact that adding mediocre alternatives did not significantly enhance optimism, even though dud alternatives did. The modification would assume that people's strategies for judging a residual are best represented as some compromise between averaging and adding. From the perspective of this compromise account, one would expect that the inclusion of dud players would cause a large decrease in the *averaged* strength of the residual but only a small increase in the *summed* strength of the residual. The net result, assuming a compromise between averaging and adding, would be a sizable decrease in the perceived strength of the residual and thus an increase in optimism about the focal outcome (as observed in the experiment). However, the inclusion of mediocre players would cause a moderate decrease in the averaged strength of the residual but also a moderate increase in the summed strength of the overall residual. The net result, assuming the averaging–adding compromise, would be no significant change in the perceived strength of the residual or optimism about the focal outcome (as we observed).

Hence, although Experiment 5 rules out a pure averaged-residual account, it does not decisively distinguish between the contrast account and a compromise version of the averaged-residual account. In the General Discussion section, however, we describe some rationales based on findings from other research for why the compromise account would seem to be a somewhat less plausible explanation for the results in the raffle ticket paradigm than is the contrast account.

Experiment 6

We assumed that many, if not most, of our research participants were well aware that adding a very weak player (i.e., a dud) to a raffle will decrease the likelihood of the focal player's winning. Nevertheless, we also assumed that under many conditions in which a person makes a likelihood judgment, this awareness about the normative influence of duds has little impact. There may be a variety of task factors that can influence the extent to which this awareness about the normative influence of duds actually manifests itself in patterns of likelihood responses. The results from Experiments 1–3 indicate that the response type (nonnumeric vs. numeric) is one such factor. Experiment 6 investigated this factor within the raffle paradigm. In this experiment, we also investigated another factor that we assumed might influence the extent to which duds have normative rather than nonnormative effects on likelihood judgments—time pressure.

We expected that time pressure would significantly hamper participants' ability and, perhaps, motivation to take a normative, rule-based approach to calculating the likelihood of winning a raffle. Hence, under time pressure, participants' awareness that weak players in a raffle can decrease their own likelihood of winning would have a reduced impact on likelihood responses. Time pressure would not, however, reduce the role of heuristic comparisons between the focal outcome and individual alternatives (i.e., the comparisons that underlie the dud-alternative effect). Therefore, the net result of time pressure would be an increase in the magnitude of dud-alternative effects. A similar experiment testing the influence of time pressure on the comparisons mediating alternative-outcomes effects found that such effects increase in magnitude under time pressure (Windschitl et al., 2003).

Whereas participants in Experiment 5 were explicitly encouraged to provide their gut-level impressions regarding their chances of winning, all participants in Experiment 6 received instructions that emphasized objectively accurate responding. They then provided either numeric or nonnumeric likelihood judgments regarding 20 raffles, and they provided those responses either at their own pace or under severe time pressure. We expected that although the instructions in this experiment would cause participants to be generally motivated to take a rule-based approach to determining their likelihood of winning, their tendency to do so—including their tendency to assess the implications of duds in a normative way—would be reduced by the nonnumeric response scale and by time pressure.

Method

Participants and design. The participants were 193 students from Elementary Psychology classes at the University of Iowa. The design was a 2 (response type: numeric or nonnumeric) × 2 (pace: time pressured or not) × 2 (raffle type: baseline or dud-present) mixed-model design with the first two factors as between participants.

Procedure. All raffles were presented on computer. As the following excerpt illustrates, the initial instructions emphasized careful and objectively correct responding:

> For this experiment, it is critical that participants attempt to give the most objectively correct or appropriate answer that they can for each raffle they see. This requires that participants remain motivated and devote high concentration for each and every raffle.

The instructions also told participants that they would be asked later whether they had been able to maintain their maximum effort for all the raffles (all but one responded "yes"). Before beginning the experiment, they were also given an opportunity to inform the experimenter whether they thought they would be unable to make a full effort on all of the raffles (none of the participants took this opportunity).

After reading these initial instructions, participants saw an example of the response scale, which differed depending on condition. Participants in the numeric-scale condition saw a row of 21 adjacent-response buttons that were labeled in 5% increments from 0% to 100%. Participants in the nonnumeric-scale condition saw a line invisibly partitioned into 21 adjacent-response buttons. The endpoints of this scale were labeled *extremely unlikely* and *extremely likely*. If participants were in the time-

pressure condition, they then read additional instructions indicating that they would have only 3 s to respond to each raffle before the picture of the raffle would disappear, at which time they would be prompted to respond immediately.

Finally, the participants started seeing and responding to the raffles. Each raffle was accompanied by the statement "Please estimate your probability of winning" and either a numeric or nonnumeric scale. After responding to a given raffle (whether under time pressure or not), participants initiated the presentation of the next raffle by clicking a "ready" button.

The raffles. Participants responded to 29 raffles—2 of which were unidentified practice raffles, 7 of which were filler raffles, and 20 of which were critical raffles. Table 6 displays the critical raffles. The critical raffles constituted 10 pairs, with each pair containing one baseline raffle and one dud-present raffle constructed similarly to those in Experiment 5. The order in which the raffles were presented was random with the constraint that if a member of a raffle pair appeared within the first 10 critical slots, the other member of the pair appeared within the last 10 slots. The first and last 10 critical slots were always separated by at least 1 filler raffle, thus ensuring that 2 raffles from the same pair never appeared in consecutive slots.

Table 6
Ticket Distributions and Objective Probabilities of Winning in Experiment 6 as a Function of Raffle Pair and Type

Raffle pair and type	Number of tickets held: "You"	Other players				Objective probability
Pair 1						
Baseline	42	46				0.477
Dud present	42	3	3	46		0.447
Pair 2						
Baseline	14	16				0.467
Dud present	14	2	16	1		0.424
Pair 3						
Baseline	23	15				0.605
Dud present	23	15	3	2		0.535
Pair 4						
Baseline	39	31				0.557
Dud present	39	31	2	3		0.520
Pair 5						
Baseline	20	16				0.556
Dud present	20	2	16	1		0.513
Pair 6						
Baseline	17	7				0.708
Dud present	17	1	7	1	2	0.607
Pair 7						
Baseline	45	53				0.459
Dud present	45	3	4	53	3	0.417
Pair 8						
Baseline	41	36				0.532
Dud present	41	36	2	2	1	0.500
Pair 9						
Baseline	15	24				0.385
Dud present	15	2	24	1	1	0.349
Pair 10						
Baseline	36	28				0.563
Dud present	36	3	3	4	28	0.486
Overall						
Baseline						0.531
Dud present						0.480

Results and Discussion

For ease of interpretation, we coded the likelihood responses on both the numeric and nonnumeric scales (which both had 21 points) from 0 to 100. We calculated, for each participant, the average response for the baseline raffles and the average response for the dud-present raffles. We also calculated a dud-alternative-effect index for each participant by subtracting his or her average response to the baseline raffles from his or her average response to the dud-present raffles. Consequently, a positive number on this index indicates that the participant gave higher responses when duds were present rather than absent (a nonnormative pattern), and a negative number indicates that the participant gave lower responses when the duds were present rather than absent (a normative pattern). Table 7 displays the average responses for the baseline and dud-present raffles as a function of response type and pace. Figure 2 represents the average dud-alternative effect as a function of those factors.

Our main predictions focused on whether the magnitude and/or direction of the dud-alternative effect would vary as a function of response type and pace. Therefore, we submitted the values from the dud-alternative-effect index to an ANOVA with response type (numeric or nonnumeric) and pace (time pressured or not) as factors.[7] As we expected, the response-type main effect was significant, $F(1, 189) = 11.78$, $p < .01$. The mean dud-alternative index was positive and significantly different from 0 in the nonnumeric-response condition ($M = 2.57$, $SD = 11.06$, $p < .05$), indicating that adding duds tended to inflate participants' nonnumeric estimates of their likelihood of winning a raffle. The mean index was negative and significantly different from 0 in the numeric-response condition ($M = -2.14$, $SD = 7.91$, $p < .01$), indicating that adding duds tended to deflate numeric estimates. Also as expected, the pace main effect was significant, $F(1, 189) = 4.97$, $p < .05$. The mean dud-alternative index was positive in the time-pressure condition ($M = 1.79$, $SD = 10.62$, $p = .11$) but negative in the no-time-pressure condition ($M = -1.25$, $SD = 8.93$, $p = .17$), although neither value was significantly different from 0. Finally, the Response Type × Pace interaction was not significant ($F < 1$).

Inspection of Figure 2 helps clarify the nature of the two main effects. It is interesting to note that for participants who responded on a numeric scale and were not under time pressure, the presence of duds in a raffle significantly reduced their optimism about winning; their mean dud-alternative index was significantly below 0 ($M = -3.80$, $SD = 6.49$, $p < .01$), suggesting that they had a general awareness that duds can lower a focal outcome's likelihood.[8] On the other hand, for participants who responded on a nonnumeric scale and were under time pressure, the presence of duds in a raffle significantly increased their optimism; their mean dud-alternative index was significantly above 0 ($M = 3.99$, $SD =$

[7] The results of a repeated-measures ANOVA that treats raffle pair (1–10) and raffle type (baseline or dud-present) as separate factors along with response type and pace yield conclusions identical to those reported in the text.

[8] Within this group of participants, the direction of this effect was generally consistent across the pairs of raffles. Specifically, for 8 of the 10 raffle pairs, the mean response to the baseline version was higher (at least directionally) than the mean response to the dud-present version.

11.79, $p < .05$).[9] In the other two cells of our design (numeric with time pressure or nonnumeric with no time pressure), the influence of duds was nonsignificant (i.e., the dud-alternative index was not significantly different from 0).

The pattern of results indicates that the duds in this experiment had notably different influences on likelihood judgments depending on both of the task factors, specifically response type and the presence or absence of time pressure. We suspect that a participant's treatment of duds in this experiment was closely tied to the extent to which he or she took a deliberative rule-based approach to judging likelihood. More specifically, although the instructions in this experiment likely caused participants to be generally motivated to take a rule-based approach to determining their likelihood of winning, we suspect that their tendency to do so—and therefore their tendency to respond to the presence of duds in a normative way—was reduced by the nonnumeric response scale and by time pressure. Support for this idea comes from additional analyses involving correlational data. For each participant, we calculated the correlation between his or her likelihood responses across the 20 critical raffles and the objective rule-based probability of winning across those raffles. These correlations were then transformed using a Fisher transformation and submitted to an ANOVA with response type and pace as between-participant factors. This analysis produced a significant response-type main effect, $F(1, 189) = 14.32$, $p < .01$, a significant pace main effect, $F(1, 189) = 18.98$, $p < .01$, and a significant interaction, $F(1, 189) = 5.56$, $p < .05$.[10] Likelihood judgments were most closely correlated with objective probabilities among participants who provided numeric responses under no time pressure (mean $r = .85$).[11] This is also the group that responded in a normative fashion to the presence of duds. Likelihood judgments were least correlated with objective probabilities among participants who provided nonnumeric responses under time pressure (mean $r = .64$). This is also the group that exhibited a nonnormative dud-alternative effect. The correlations among the numeric-with-time-pressure group (mean $r = .69$) and nonnumeric-with-no-time-pressure group (mean $r = .71$) were intermediate. These were also the two groups in our design that showed no significant effects for the presence or absence of duds. In sum, it appears that the participants' treatment of duds was, in fact, closely tied to the extent to which they took a rule-based approach to judging likelihood, which was substantially influenced by both time pressure and response mode.

Table 7
Mean Likelihood Responses for the Baseline and Dud-Present Raffles as a Function of Response Type and Pace

	Pace			
	No time pressure		Time pressure	
Response type	*M*	*SD*	*M*	*SD*
Numeric condition				
Baseline	55.62	6.40	55.79	6.43
Dud present	51.83	9.07	55.37	10.09
Nonnumeric condition				
Baseline	59.72	9.21	57.04	8.29
Dud present	60.96	14.07	61.03	12.75

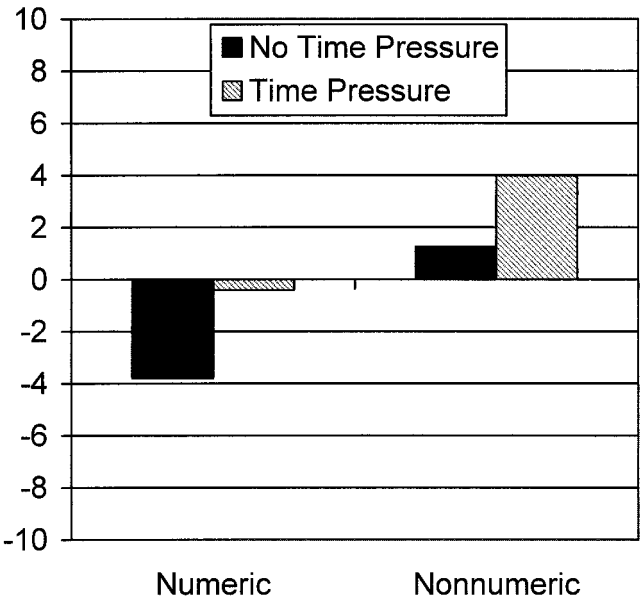

Figure 2. The dud-alternative-effect index in Experiment 6. Positive numbers indicate that the presence of duds increased participants' likelihood judgments about winning. Negative numbers indicate that the presence of duds decreased participants' likelihood judgments about winning.

General Discussion

The experiments in this article provide novel information about how people's evaluations of evidence for multiple possible alternative outcomes ultimately shape their likelihood judgments. A normative model of subjective probability suggests that when weak alternatives (duds) are added to a list of alternative outcomes, the judged probability of a focal outcome should decrease or perhaps remain unchanged (if the duds are completely implausible) but should never increase. However, our studies demonstrated that adding duds to a list of alternative outcomes often does lead to an increase in the judged likelihood of a focal outcome.

Support theory's original formulation, which can successfully model a variety of likelihood judgment phenomena, does not anticipate our findings regarding the dud-alternative effect (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994). Support theory, like the averaged-residual account, assumes that people evaluate the evidence in the residual as a whole—that is, they judge the overall support for all alternatives to the focal hypothesis (see Brenner & Koehler, 1999; Koehler, Brenner, & Tversky, 1997). Although a key claim of support theory is that the perceived support for the full residual is often less than the sum of the perceived support for the individual hypotheses in

[9] Within this group of participants, the direction of this effect was generally consistent across the pairs of raffles. Specifically, for 9 of the 10 raffle pairs, the mean response to the dud-present version was higher (at least directionally) than the mean response to the baseline version.

[10] We also computed gamma correlations for each participant and submitted those to the same type of ANOVA. This ANOVA yielded the same results as did the ANOVA reported for the Pearson correlations.

[11] The reported means were transformed back to standard correlation coefficients (reversing the Fisher transformation).

the residual (*subadditivity*), the theory does not further specify how support assessments for individual alternative hypotheses are treated or integrated when a person is gauging the overall support for the residual (although see discussion by Brenner & Koehler, 1999; Koehler et al., 1997). The common attribution for subadditivity assumes that evidence regarding a hypothesis is least salient (and perhaps forgotten) when that hypothesis is implicitly packed within a residual, more salient when the hypothesis is explicitly unpacked within the residual, and most salient when the hypothesis is directly evaluated. From this perspective, one might have predicted that the addition of a dud would make salient an additional set of evidence that supports the residual, thereby lowering the judged likelihood of the focal hypothesis. On the other hand, one might have proposed that the addition of a dud would make salient an exceptionally weak set of information that would decrease the perceived strength of the overall residual, thereby raising the judged likelihood of the focal hypothesis. Given that support theory could have supported opposing predictions, it appears that the theory was agnostic on the issue of how duds would influence focal likelihood judgments.

From this agnostic position, however, it is possible that support theory could be extended to accommodate the dud-alternative effect. In fact, the averaged-residual account borrows all of the same assumptions as support theory but simply adds a specific proposal as to how evidence for individual hypotheses is integrated when judging the strength of the residual. The strong version of that account could not account for the results of Experiment 5 involving the mediocre-present condition, but the version of the account that assumes a compromise between averaging and adding could account for those effects as well as the basic dud-alternative effect (e.g., from Experiment 1). Perhaps a bigger challenge for support theory would be to accommodate the fact that the magnitude of the dud-alternative effect shifts as a function of response mode (Experiments 3 and 6) and time pressure (Experiment 6).

A novel account on which we focused was the contrast account. This account does not assume that participants attempt to form an aggregated representation of evidence for the hypotheses in the residual but instead that they compare the evidence for the focal outcome to evidence for the individual alternative outcomes. When duds are added to the local context (i.e., the list of alternatives), the focal option compares very favorably to a greater number of individual alternatives, and therefore the perceived strength of the evidence for the focal option is enhanced.

Although support theory, with a compromise version of the averaged-residual account, could explain the basic dud-alternative effect, there are reasons to conclude that the mechanisms described by the contrast account played an important role if not the key role in the dud-alternative effects detected in our experiments. First, Experiment 4 provided evidence that the presence of duds can influence the judged strength of the focal option. For example, when asked "Compared to all shoe brands on the U.S. market, how attractive is Nike to the typical U.S. teenage boy?," participants gave higher responses when they had just read an option list that included duds (Etonic and Keds) than when they had just read one that excluded those duds. This contrast effect was not predicted by the averaged-residual account (neither the pure nor compromise version), but it is quite consistent with the key processes proposed within the contrast account.

A second reason to assume the mechanisms of the contrast account played a key role in the dud-alternative effects concerns research by Rottenstreich et al. (1999). In that research, participants made likelihood judgments about disjunctive and component events for which the relevant evidence was either similarity based ("How likely is this described person a journalist or realtor rather than an insurance salesperson?") or frequency based ("How likely is a randomly selected American a journalist or realtor rather than an insurance salesperson?"). Using a support-theory framework, Rottenstreich et al. computed implicit support assessments underlying the likelihood judgments. They found that for similarity-based likelihood judgments, the perceived support for the disjunction was often less than the perceived support for one of the components when individually judged. This finding, which is consistent with an averaging account, is relevant to Experiments 1–3, given that many of the judgments in those experiments could be based on similarity assessments. However, Rottenstreich et al. also found that for frequency-based likelihood judgments, the perceived support for the disjunction was typically not less than the perceived support for a single component. This finding, which suggests that support for components of a residual was not averaged, is relevant to our raffle paradigm given that participants' likelihood judgments were based entirely on frequency information. Hence, the findings of Rottenstreich et al. suggest that although averaging may have contributed to the dud-alternative effects detected within the question paradigm (Experiments 1–3), averaging was unlikely to have played a key role in the effects detected within the raffle paradigm (Experiments 5 and 6). The contrast account would seem to be applicable to the effects in both paradigms.

A third and final point supporting the role of the contrast account is related to prior research that used the raffle paradigm to investigate the alternative-outcomes effect. As described in our introduction, Windschitl and Young (2001) demonstrated how manipulations to the distribution of tickets across players in a residual (which hold the average and sum of the tickets in the residual constant) affect participants' optimism about winning. Neither averaging nor adding (nor a combination of the two) can account for these alternative-outcomes effects, and therefore, it seems implausible that averaging would be the key process mediating dud-alternative effects in Experiments 5 and 6. The contrast account discussed in this article, however, is in fact compatible with the account that has been offered to explain the alternative-outcomes effect (Windschitl & Wells, 1998; Windschitl & Young, 2001); alternative-outcomes effects occur because the comparison between the focal and the strongest alternative has a disproportionate influence relative to other comparisons, and the dud-alternative effect occurs because increasing the number and magnitude of favorable pairwise comparisons (i.e., favorable for the focal outcome) can influence the perceived strength and likelihood of the focal outcome.

Detecting Dud-Alternative Effects

The general idea that duds might influence the subjective strength of evidence for a focal option via contrast effects is well grounded in the numerous studies demonstrating how contextual stimuli within a local context can influence target judgments (see discussions by Eiser, 1990; Helson, 1964; Higgins & Lurie, 1983; Kahneman & Miller, 1986; Parducci, 1965; Wedell, 1994). Although contrast effects are prevalent in many forms and often readily detectable, the contrast effects of the sort studied here are perhaps difficult to discover because their influence is typically offset by a different type of process. People are generally aware

that for making a likelihood judgment (as opposed to other judgments such as strength judgments), evidence that supports any alternative—even a weak one—should reduce the likelihood of the focal option (the complementarity rule).[12] Hence, adding a weak alternative might increase the perceived strength of a focal option via contrast but simultaneously lower its judged likelihood because of likelihood-estimation processes that are at least partially sensitive to the complementarity rule.

We suspected, therefore, that dud-alternative effects would be most likely to be detected when: (a) The duds are viewed as nearly or completely implausible by the respondents and (b) The judgment task does not precipitate participants' concerns with the complementarity issue, or task factors inhibit rule-based processing, thereby hampering participants' use of the complementarity rule. To minimize participants' concerns with the complementarity issue, our key dependent measure for testing the dud-alternative effect was a nonnumeric likelihood scale; previous research has suggested that standard numeric measures tend to enhance participants' concerns with normative probability rules (see Windschitl, 2000; Windschitl & Wells, 1996, 1998). Indeed, the dud-alternative effects were strong and reliable on nonnumeric measures (Experiments 1, 2, 3, 5, and 6), but as we predicted, the effects were nonsignificant or reversed for participants who were asked to respond on a 100-point probability scale (Experiments 2, 3, and 6). To ensure that the effect detected by the nonnumeric measures reflected consequential changes in internal perceptions of certainty rather than merely changes in participants' use of the response scale, we also tested and found that betting measures showed the same sensitivity to the dud alternatives as did the nonnumeric likelihood measures (Experiments 2 and 3). This pattern of findings for the dud-alternative effect across nonnumeric, numeric, and betting measures is the same as that found for the alternative-outcomes effect (see Windschitl et al., 2003; Windschitl & Wells, 1998; Windschitl & Young, 2001). In sum, although dud-alternative effects are detectable only under a particular set of conditions (a and b above), the mechanisms driving the effect may nevertheless influence a wide variety of likelihood judgments in ways that are not easily detected because of the influence of offsetting processes.

Examples of real-world judgments for which the mechanisms driving the dud-alternative effect may be particularly relevant are confidence judgments that eyewitnesses make when viewing line-ups. Researchers have already noted hazards of using line-up distractors or foils that bear little resemblance to the description of the culprit (see e.g., Gonzalez, Davis, & Ellsworth, 1995; Wells et al., 1998). When an innocent suspect who happens to fit the culprit's description is accompanied in a lineup by only ill-fitting distractors, the number of false identifications of the suspect (and confidence in those identifications) can increase dramatically relative to when the suspect is accompanied by distractors that do fit the general description of the culprit (see Wells, Rydell, & Seelau, 1993). Our work suggests that merely adding ill-fitting distractors (duds) to a lineup that already contains well-fitting distractors could increase an eyewitness's confidence about whomever they pick from the lineup. Relatedly, the inclusion of the ill-fitting distractors could influence eyewitnesses' judgments of how well the selected person matches their memory of the culprit (just as the duds in our Experiment 4 influenced strength judgments).

Two Types of Processes in Likelihood Judgment

We did not set out to test an overall theory of likelihood judgment, but our findings, coupled with findings regarding the alternative-outcomes effect, provide some grounds for speculation about how an overall theory might account for those effects as well as effects explained by support theory. First, one should consider two extremes: Sometimes people make likelihood judgments almost immediately and without any deliberative thought or effort, whereas sometimes people make likelihood judgments after careful and deliberative thought in an effort to provide the best probability value. At the former extreme, people make no attempt to sum or otherwise aggregate the evidence for the alternative outcomes; instead, pairwise comparisons between the focal and individual alternatives play the key role in shaping likelihood judgments. The alternative-outcomes effect occurs because the comparison between the focal and the strongest alternative, relative to other comparisons, plays a disproportionate role in shaping the perceived strength and likelihood of the focal outcome. The dud-alternative effect occurs because the perceived strength and likelihood of the focal outcome is enhanced by very favorable pairwise comparisons within the local context. At the latter extreme, people attempt to judge likelihood by assessing the balance of evidence for and against the focal outcome, as suggested by support theory. To do this, they consider evidence for all individual alternatives and then compute the overall support for the full set of these alternatives in the residual. This computation approximates an additive model but tends to fall short (subadditivity) because the memory and salience issues underlying subadditivity are not easily avoided, even though people at this extreme are motivated to think effortfully and to give their best estimate of probability.[13]

Although we have described two extremes, we suspect, in fact, that most likelihood judgments fall somewhere between those extremes. In other words, both the pairwise comparisons and additive-aggregation processes can contribute to a given judgment; various task factors (e.g., accountability, distraction) and perhaps individual differences (e.g., expertise) are likely to determine the relative influence of the two types of processes. The present research and related research on the alternative-outcomes effect (Windschitl et al., 2003; Windschitl & Wells, 1998) suggest that numeric probability questions and the absence of time pressure can enhance the relative influence of additive-aggregation processes, whereas nonnumeric questions and the presence of time pressure allow for a greater influence of pairwise comparison processes.

Further definition of this two-process framework requires additional research involving likelihood judgments made in cases for

[12] Research has suggested that people are far from perfect when it comes to keeping their likelihood judgments in conformance with this complementarity rule. However, these findings do not mean that people are oblivious to the complementarity issue or that their likelihood judgments show absolutely no patterns consistent with a complementarity constraint.

[13] Although we are speculating that support theory is applicable for describing a situation in which people are taking an effortful approach to estimating a probability, we caution that support theory is agnostic about exactly how the support for a residual is computed; effortful aggregation has not been stipulated in descriptions of support theory.

which there are multiple possible alternatives; the distinctions between the two types of processes (pairwise comparisons and additive-aggregation) disappear in binary cases for which there is only one nonfocal alternative and hence no need to aggregate evidence for a residual. Such research on multialternative cases seems worthwhile, as this proposed framework offers some promise for achieving a fuller understanding, beyond that revealed by support theory, of how people go about thinking of and using evidence when judging likelihood under a variety of conditions.

References

Ariely, D., & Wallsten, T. S. (1995). Seeking subjective dominance in multidimensional space: An explanation of the asymmetric dominance effect. *Organizational Behavior and Human Decision Processes, 63,* 223–232.

Bar-Hillel, M., & Neter, E. (1993). How alike is it versus how likely is it: A disjunction fallacy in probability judgments. *Journal of Personality and Social Psychology, 65,* 1119–1131.

Brenner, L. A., & Koehler, D. J. (1999). Subjective probability of disjunctive hypotheses: Local weight models for decomposition of evidential support. *Cognitive Psychology, 38,* 16–47.

Campbell, D. T., Lewis, N. A., & Hunt, W. A. (1958). Context effects with judgmental language that is absolute, extensive, and extra-experimentally anchored. *Journal of Experimental Psychology, 55,* 220–228.

Carlson, B. W., & Yates, J. F. (1989). Disjunction errors in qualitative likelihood judgment. *Organizational Behavior and Human Decision Processes, 44,* 368–379.

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

Eiser, J. R. (1990). *Social judgment.* Milton Keynes, England: Open University Press.

Gonzalez, R., Davis, J., & Ellsworth, P. C. (1995). Who should stand next to the suspect? Problems in the assessment of lineup fairness. *Journal of Applied Psychology, 80,* 525–531.

Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior.* New York: Harper & Row.

Herr, P. M. (1986). Consequences of priming: Judgment and behavior. *Journal of Personality and Social Psychology, 51,* 1106–1115.

Higgins, E. T., & Lurie, L. (1983). Context, categorization, and recall: The "change-of-standard" effect. *Cognitive Psychology, 15,* 525–547.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology, 24,* 1–55.

Huber, J., Payne, J. W., & Puto, C. (1982). Adding asymmetrically dominated alternatives: Violations of regularity and the similarity hypothesis. *Journal of Consumer Research, 9,* 90–98.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

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.

Koehler, D. J., Brenner, L. A., & Tversky, A. (1997). The enhancement effect in probability judgment. *Journal of Behavioral Decision Making, 10,* 293–313.

Krantz, D. L., & Campbell, D. T. (1961). Separating perceptual and linguistic effects of context shifts upon absolute judgments. *Journal of Experimental Psychology, 62,* 35–42.

Manis, M. (1967). Context effects in communication. *Journal of Personality and Social Psychology, 5,* 326–334.

Manis, M., & Armstrong, G. W. (1971). More context effects in communication. *Journal of Experimental Social Psychology, 7,* 381–388.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Monographs, 77*(2, Whole No. 565).

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., Brenner, L., & Sood, S. (1999). Similarity between hypotheses and evidence. *Cognitive Psychology, 38,* 110–128.

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

Simonson, I., & Tversky, A. (1992). Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research, 29,* 281–295.

Simpson, D. D., & Ostrom, T. M. (1976). Contrast effects in impression formation. *Journal of Personality and Social Psychology, 34,* 625–629.

Stevens, S. S. (1958). Adaptation-level vs. the relativity of judgment. *The American Journal of Psychology, 71,* 633–646.

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 and 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.

Upshaw, H. S. (1969). Stimulus range and the judgmental unit. *Journal of Experimental Social Psychology, 5,* 1–11.

Wedell, D. H. (1991). Distinguishing among models of contextually induced preference reversals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 4,* 767–778.

Wedell, D. H. (1994). Contextual contrast in evaluative judgments: A test of pre- versus postintegration models of contrast. *Journal of Personality and Social Psychology, 66,* 1007–1019.

Wedell, D. H., & Pettibone, J. C. (1996). Using judgments to understand decoy effects in choice. *Organizational Behavior and Human Decision Processes, 67,* 326–344.

Wells, G. L., Rydell, S. M., & Seelau, E. P. (1993). The selection of distractors for eyewitness lineups. *Journal of Applied Psychology, 78,* 835–844.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior, 22,* 1–39.

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., Krizan, Z., & Flugstad, A. R. (2003). *Likelihood judgment in multi-alternative cases.* Manuscript in preparation, University of Iowa.

Windschitl, P. D., Martin, R., & Flugstad, A. R. (2002). Context and the interpretation of likelihood information: The role of intergroup comparisons on perceived vulnerability. *Journal of Personality and Social Psychology, 82,* 742–755.

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.

Windschitl, P. D., & Young, M. E. (2001). The influence of alternative outcomes on gut-level perceptions of certainty. *Organizational Behavior and Human Decision Processes, 85,* 109–134.

Windschitl, P. D., Young, M. E., & Jenson, M. E. (2002). Likelihood judgment based on previously observed outcomes: The alternative-outcomes effect in a learning paradigm. *Memory & Cognition, 30,* 469–477.

(*Appendixes follow*)

Appendix A

Questions Used in Experiment 1

Below are dud-present versions of the questions that were first used in Experiment 1. The duds that were removed to create dud-absent versions of these questions are shown in italics. Asterisks identify the focal options, which were the same for both versions of each question. Each question asked respondents to indicate how likely it was that the focal option was the most frequently selected option.

In a recent survey, a random sample of U.S. high school students was asked: "Which of the following candies is the one you most often buy?"

Nut Rolls, *Snickers, *Pay Day,* M&Ms

In a recent survey, a random sample of U.S. high school students was asked: "Which of the following ice cream flavors is most attractive to you?"

*chocolate fudge, *orange,* peanut butter fudge, *raspberry*

In a recent survey, a random sample of U.S. female college students was asked: "Which of the following actors do you find most attractive?"

Jason Alexander (George from Seinfeld), Bruce Willis (movies, e.g., *Die Hard*), *Jackie Chan (movies, e.g., Rush Hour),* *Brad Pitt (movie, e.g., *Seven*)

In a recent survey, a random sample of U.S. boys aged 13–17 was asked: "Which of the following athletes would you most like to meet?"

Tiger Woods, *A.C. Green,* *Michael Jordan, *Lee Trevino*

In a recent survey, a random sample of U.S. female college students was asked: "Which of the following music groups do you prefer the most?"

Kansas, *Dave Matthews Band, *Milli Vanilli,* Phish

In a recent survey, a random sample of U.S. females aged 19–25 was asked: "Which of the following sitcoms do you prefer the most?"

Seinfeld, *MASH, Saved by the Bell,* *Friends

In a recent survey, a random sample of U.S. children aged 7–10 was asked: "Which of the following is your favorite type of food for dinner?"

*Pizza, Hamburger, *Eggplant Parmesan, Grilled fish*

In a recent survey, a random sample of U.S. adults was asked: "Which of the following types of ethnic restaurants do you eat at the most?"

*Mexican, *Caribbean, Moroccan,* Italian

In a recent survey, a random sample of U.S. teenage boys was asked: "Which of the following sports do you most prefer to watch?"

men's gymnastics, men's bowling, Major League Baseball, *NBA Basketball

In a recent survey, a random sample of U.S. teenage boys was asked: "Which of the following shoe brands do you most prefer?"

Keds, Etonic, Adidas, *Nike

Appendix B

New Questions Used in Experiment 3

Below are dud-present versions of the "new" questions used in Experiment 3. The duds are in italics, and the asterisks identify the focal options.

One, and only one, of the following cities is the capital of the Alberta province of Canada.

Seattle, *Regina, Edmonton, *Detroit*

One, and only one, of the following cities lies to the south of the equator.

Cincinnati, Ohio; *Nairobi, Kenya; *Moscow, Russia;* Calcutta, India

One, and only one, of the following newspapers has the largest daily circulation of any newspaper in the United States.

The Los Angeles Times, *The Colorado Springs Gazette, The Des Moines Register,* *The New York Times

One, and only one, of the following is a country located in Europe.

Philippine Islands, New Zealand, Uzbekistan, *Slovakia

One, and only one, of the following is the most popular form of travel in European countries.

hot air balloon, automobile, *train/rail, *helicopter*

One, and only one, of the following is the most frequently appearing letter in English words.

E, *Y, U,* *A

One, and only one, of the following is the most popular choice of color for a new automobile.

Lime, Blue, *Orange,* *Red

One, and only one, of the following is the most frequently occurring traffic infractions actually resulting in a ticket.

Failure to stop at a stop sign, *Speeding, *Failure to use turn signal, Too many occupants in vehicle*

One, and only one, of the following cities is the most popular vacation destination in the United States.

Wheeling, West Virginia; *Los Angeles, California; *Little Rock, Arkansas;* Miami, Florida

One, and only one, of the following is the greatest selling male solo vocalist in history.

Michael Jackson, *John Tesh,* *Elvis Presley, *Kenny Loggins*

One, and only one, of the following cities hosted the 1984 Summer Olympic Games.

*Los Angeles, California; *Lexington, Kentucky;* Seoul, South Korea; *Kansas City, Missouri*

One, and only one, of the following is the highest paid actor in history.

Ice Cube, McCauley Caulkin, Arnold Schwarzenegger, *Tom Cruise

One, and only one, of the following cities lies on the Mediterranean Sea.

Cape Town, South Africa; Cairo, Egypt; *Naples, Italy; *Jakarta, Indonesia*

One, and only one, of the following is the most popular recreational activities in Yosemite National Park.

mountain biking, *water skiing, ice climbing,* *hiking

One, and only one, of the following geographic regions has the coldest average temperature in the world.

Siberia, *Portugal, Italy,* *Antarctica

One, and only one, of the following people grew up in England.

*Elton John, *George W. Bush,* Sean Connery, *Jennifer Lopez*

The 2001 Tour De France was primarily located in France, but at one point during the race it crosses over into another country.

Italy, *Pakistan,* *Belgium, *Egypt*

One, and only one, of the following is the most common source of allergies in late adulthood.

*domestic animals, *blackberry jam,* pollen and mold, *cinnamon*

Received March 30, 2003
Revision received July 9, 2003
Accepted July 11, 2003 ■