

Contingent Approaches to Making Likelihood Judgments about Polychotomous Cases: The Influence of Task Factors

PAUL D. WINDSCHITL* and ZLATAN KRIZAN
University of Iowa, USA

ABSTRACT

Two experiments tested the influence of three task factors on respondents' tendency to use normative, heuristic, and random approaches to making likelihood judgments about polychotomous cases (i.e., cases in which there is more than one alternative to a focal hypothesis). Participants estimated their likelihood of winning hypothetical raffles in which they and other players held various numbers of tickets. Responding on non-numeric scales (vs. numeric ones) and responding under time pressure (vs. self-paced) increased participants' use of a comparison-heuristic approach, resulting in non-normative judgment patterns. A manipulation of evidence representation (whether ticket quantities were represented by numbers or more graphically by bars) did not have reliably detectable effects on processing approaches to likelihood judgment. The authors discuss the implications of these findings for the further development of likelihood judgment theories, and they discuss parallels between contingent processing in choice and contingent processing in likelihood judgment. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS likelihood judgment; probability; heuristics; contingent processes; alternative-outcomes effect

INTRODUCTION

Research on contingent decision making has shown that people typically have a variety of options for processing attribute information regarding choice alternatives (Einhorn & Hogarth, 1981; Payne, 1982; Payne, Bettman, & Johnson, 1992, 1993; Svenson, 1979). *Lexographic, elimination-by-aspects, and weighted additive rule* are just some of the terms that have been developed to describe various strategies a decision maker can use to process information about alternatives and arrive at a decision (e.g., Tversky, 1972). Research findings also suggest that people's selection of a processing strategy is often quite flexible and adaptive. For example, people who are adequately motivated will use cognitively demanding strategies that maximize

*Correspondence to: Paul D. Windschitl, Department of Psychology, University of Iowa, Iowa City, Iowa, 52242, USA.
E-mail: paul-windschitl@uiowa.edu

Contract/grant sponsor: National Science Foundation; contract/grant number: SES 99-11245.

accuracy when the task factors afford this level of cognitive effort, but they will tend to shift toward cognitively simpler strategies as other task demands or opportunity costs increase (e.g., Payne et al., 1993; Payne, Bettman, & Luce, 1996; Russo & Doshier, 1983).

Within the domain of probability judgment, there have also been major pockets of research devoted to understanding when various contingent processes dominate in the judgment-formation processes. For example, the use of base-rate and representativeness information has been shown to be a function of a variety of factors, such as prior activation of inferential rules (Ginossar & Trope, 1987) and whether the base-rate information is represented and measured in a frequentist or probabilistic format (Gigerenzer, 1991, 1996; Kahneman & Tversky, 1996).

Whereas there has been substantial research on contingent decision making and contingent processing for some forms of probability judgment, there has been relatively little work done on the contingent processes regarding a particular type of likelihood judgment—namely judgments that involve polychotomous cases. Polychotomous cases are ones in which people are judging the likelihood that a focal outcome will occur (or a focal hypothesis is true) when there is more than one alternative to that focal outcome (or hypothesis). In these cases, a respondent must consider the strength of evidence for the focal outcome but also for each of the multiple alternative outcomes. For example, when asked “What is the likelihood that in a random selection from all chemistry, biology, geology, and physics students at University X, a chemistry student would be selected?”, a respondent must consider the prevalence of not only chemistry students, but also the biology students, geology students, and physics students. Although it seems intuitively plausible that there could be various ways in which a respondent facing that question might go about processing information and arriving at an answer, existing theories of likelihood judgment, such as support theory (see Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994), do not specify how likelihood judgment processes might change as a function of task conditions that are known to vary in real world settings (e.g., time pressure).

The experiments in this paper investigated the influence of such task conditions on the processes mediating likelihood judgments in polychotomous cases. We presented people with uncertain situations involving stochastic outcomes, and the amount of evidence supporting each possible outcome was explicitly represented. More specifically, participants saw depictions of hypothetical raffles in which they and other players held various numbers of tickets (see Figure 1a). For each raffle, they estimated their likelihood of winning. In these experiments, we distinguish between 3 types of contingent processes (or approaches) for judging the likelihood of winning the raffles: a *normative approach*, *comparison-heuristic approach*, and *random approach*.

CONTINGENT APPROACHES

The normative approach

For people who have a formal education in math and statistics, the normative approach to determining one's likelihood of winning the raffle depicted in Figure 1a should be clear. Assessing the likelihood of winning would require respondents to divide the number of tickets they held by the total number of tickets in the raffle [e.g., $14/(14 + 3 + 4 + 13 + 3 + 3) = 0.35$]. This approach to judging likelihood within the raffle paradigm would (as suggested by the label “normative”) lead to an objectively correct judgment on each occasion, assuming that processing disruptions are not an issue. Depending on the number of players and tickets in the raffle, this process involving the summing of tickets across alternatives could be quite resource intensive. Also, Kahneman (2003) recently noted that additive representations of stimuli, which are required by this approach, are not readily accessible; generating a representation of the sum of a set of stimuli requires deliberative intent to do so. Indeed, we assume that the processes comprising this normative approach have characteristics in common with other forms of deliberative and rule-based processes (see, e.g., Sloman, 1996 for review).

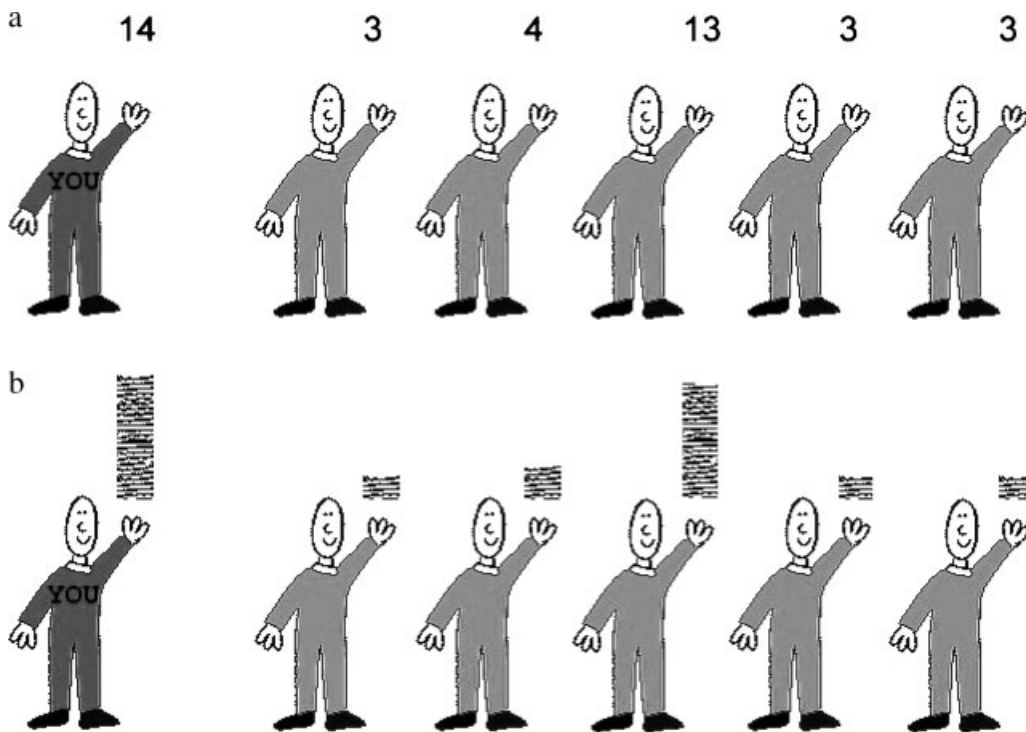


Figure 1. (a) Example of a raffle used in Experiments 1 and 2 with evidence represented numerically (b) Example of a raffle used in Experiments 1 and 2 with evidence represented graphically

The comparison-heuristic approach

Even if the normative solution for judging the likelihood of winning raffles is known by most individuals, research using this raffle-ticket paradigm and related paradigms has revealed that people often do not exclusively use this normative approach to judging likelihood (Gonzalez & Frenck-Mestre, 1993; Teigen, 2001; Windschitl & Wells, 1998; Windschitl & Young, 2001). Windschitl and Wells (1998) argued that when people are asked to provide non-numeric likelihood judgments about focal outcomes in polychotomous cases, they often rely in part on a *comparison heuristic*. More specifically, people make pairwise comparisons between the evidence for the focal outcome and the evidence relevant to each alternative outcome. The pairwise comparison involving the strongest of all the alternative outcomes has a dominating influence on a participant's intuitive level of certainty about the focal outcome—the more this comparison favors the focal outcome (or the less it favors the strongest alternative), the greater the perceived likelihood of the focal outcome. Consistent with the idea that people sometimes use a comparison heuristic, Windschitl and Wells (1998) demonstrated systematic instances of a phenomenon called the alternative-outcomes effect. In demonstrations of the effect, people's perceptions of certainty regarding a focal outcome 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. In one study, participants' intuitive certainty about winning (expressed on a non-numeric scale) was greater for a raffle in which they were said to hold 21 tickets and others held 14, 13, 15, 12, and 13 tickets than a raffle in which they were said to hold 21 tickets and others held 52, 6, 2, 2, and 5 tickets. The direction of this effect is consistent with the comparison heuristic because the comparison

between the evidence for the focal outcome and the strongest alternative outcome is more favorable for the focal outcome in the former raffle (21 vs. 15 tickets) than in the latter raffle (21 vs. 52 tickets). In this and other demonstrations of the alternative-outcomes effect (and a closely related effect called the equiprobability effect; Teigen, 2001), the observed differences in likelihood judgments have consistently fallen in a direction supporting the hypothesized comparison heuristic.

Additional evidence for the notion that alternative-outcome effects are due to pairwise heuristic comparisons comes from a study in which participants were given an opportunity to select which of two 25-ticket raffles they wanted to hold a ticket in—one raffle in which each person held only one ticket or a second raffle in which one person held seven tickets but everyone else each held one ticket (Windschitl & Wells, 1998). Most participants preferred to play in the former raffle. Moreover, their self-reports for why they picked the former raffle often indicated that they were comparing their chances of winning to each individual player; they felt less optimistic about the latter raffle because there was a player in that raffle who held many more tickets than they did.

It is important to emphasize that the comparison-heuristic approach provides answers that will generally but not perfectly track objective probabilities (see Windschitl & Wells, 1998). It would also appear to be an approach that is more spontaneous and less resource demanding than the normative approach, because the process requires only pairwise comparisons rather than the deliberative summation of tickets. Finally, unlike the normative and random approaches, only the comparison-heuristic approach produces systematic cases of alternative-outcomes effects.

The random approach

A third “approach” to judging likelihood is a random approach, in which participants, for whatever reason, respond in an apparently random fashion across the raffles they encounter. Responses from this approach would necessarily be uncorrelated with objective probabilities and would not exhibit any alternative-outcomes effects. A random approach to estimating likelihood would presumably be the least resource demanding of the three approaches.

Summary

As is frequently done in the literature on contingent decision making (see, e.g., Payne et al., 1993), we have placed the approaches we are investigating on a continuum ranging from normative to random. The mid-level approach we have identified (i.e., the comparison-heuristic approach) has a heuristic property—or an efficiency property that might make it adaptive under some task conditions. Our focus in this paper is primarily on the first two approaches. The random approach is not necessarily a true approach but rather the absence of a systematic strategy to making likelihood judgments. Nonetheless, an important question addressed in this research is whether people who cannot or do not use a normative approach because of adverse task conditions would switch to noisy responding (a random approach) or whether they would switch to another form of systematic responding (the comparison-heuristic approach).

THE MODERATORS AND PREDICTIONS

Figure 2 depicts a framework for understanding the conceptualization and predictions of this research. Because an underlying goal of this research was to explore the degree of flexibility in processing for judgments in polychotomous cases, we chose potential moderator variables that varied across real-world settings and that could be plausibly expected to influence a respondent’s processing approach. Specifically, we tested

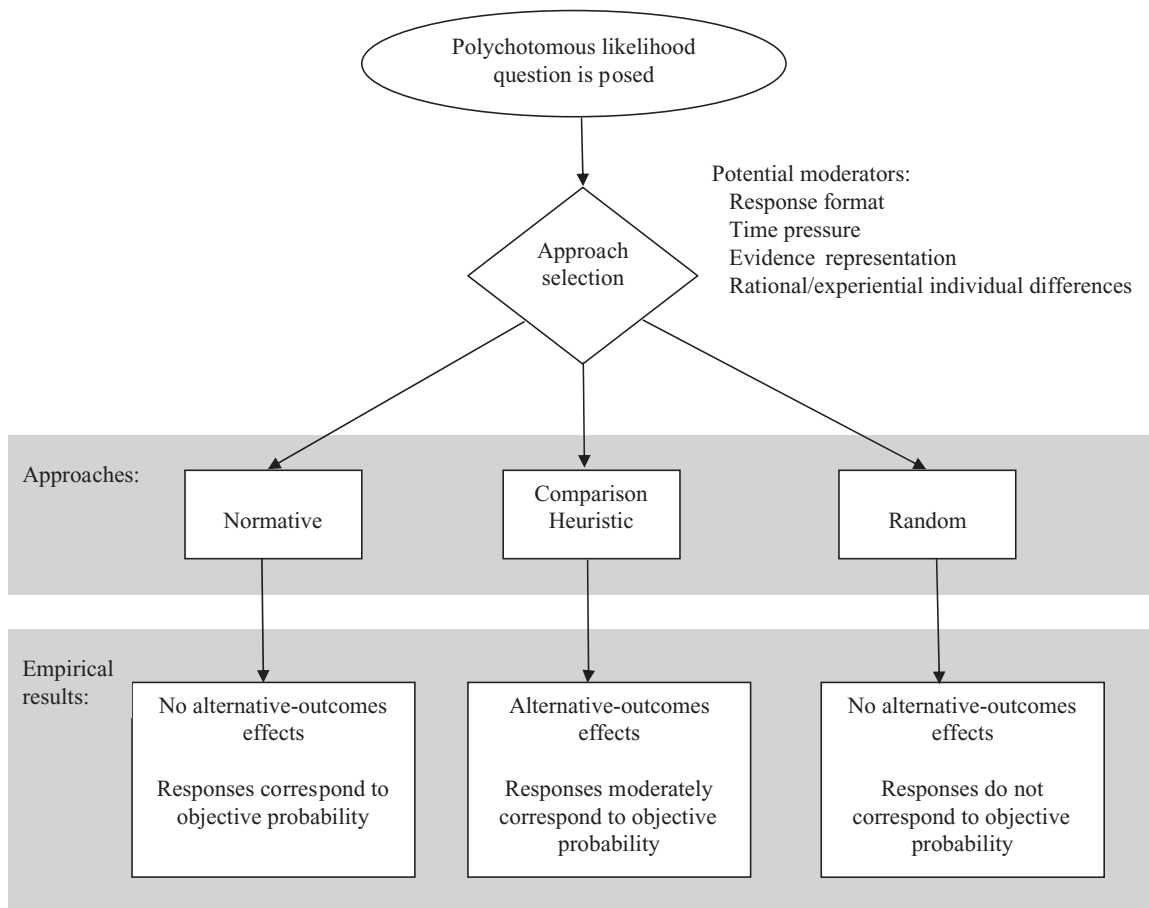


Figure 2. A conceptualization for the possible approaches, moderators, and empirical results that were examined in this research. The “Empirical results” box shows hallmarks or key results that would occur if a person were to use the specified approach. For example, a participant using the comparison-heuristic approach would produce likelihood judgments that contained alternative-outcomes effects and would be moderately correlated with objective probabilities. Although the conceptualization may appear to suggest exclusivity among the approaches, we do not assume that a participant would use only one approach for judging the likelihoods of winning raffles

whether the approaches that people take toward judging their likelihood of winning a raffle differed as a function of three manipulated variables: 1) whether participants were required to provide numeric or non-numeric estimates of likelihood, 2) whether they were self-paced or placed under time pressure, and 3) whether the quantity of tickets held by each raffle player was represented by a number (as is the case in Figure 1a) or by the height of a “ticket stack” (see Figure 1b). Experiment 1 also tested the influence of two individual difference variables that reflect people’s preferences for rule-based and intuitive thinking.

The response format factor

Unlike the other two factors being investigated, the influence of the response-format factor on the use of the comparison heuristic has been investigated in previous work. Alternative-outcomes manipulations have

tended to produce robust effects on verbal and other non-numeric measures (e.g., a measure that includes options such as “somewhat likely”) but no discernable effects on numeric probability measures (e.g., Teigen, 1988, 2001; Windschitl & Wells, 1996, 1998). When participants in a typical alternative-outcomes study are asked to respond on a numeric scale, there is a compatibility between the response scale values and the numeric values the participants would use when mentally applying a formal rule to calculate an answer (Slovic, Griffin, & Tversky, 2002). This compatibility can prompt a heightened level of deliberative and rule-based thinking, because people become aware that there is a right and a wrong answer and that they know some formal rules that can be applied to arrive at the right answer (Windschitl & Wells, 1996, 1998). However, when asked to provide a non-numeric estimate, respondents are aware that there is no precisely correct or incorrect answer, which reduces the need for deliberate and rule-based processing. Instead, participants can rely, at least in part, on other processes to make judgments—in this case, a comparison-heuristic approach.

In the present experiments, we expected to replicate the previously observed differences in the extent to which numeric and non-numeric measures detect alternative-outcomes effects (e.g., Teigen, 1988, 2001; Windschitl & Wells, 1998). Unlike previous studies that have tested this issue on a scenario-by-scenario basis, our paradigm examined patterns of responses across several raffles, which provided for more stable estimates about the extent to which numeric and non-numeric measures tend to elicit normative rather than heuristic comparison processes. We also recorded the response latencies between the onset of a given raffle and the recording of a response. Although both the numeric and non-numeric measures required only a mouse click, we expected that participants who were asked to give numeric rather than non-numeric responses would show relatively longer latencies (when not under time pressure), reflecting greater degrees of deliberative and rule-based processing mediating their responses (i.e., the use of formal probability rules).

The time pressure factor

What happens to people’s likelihood judgments and the processes that mediate them under time pressure? Generating an accurate likelihood judgment would appear to become increasingly difficult as time pressure increases. For example, the calculations required to determine the objective probability of winning the raffle in Figure 1a (summing the total number of tickets and dividing “your” number of tickets by that value) would presumably be quite difficult to do within a 4-second deadline. Hence, one possibility is that likelihood judgments degenerate in a random fashion and simply reflect greater and greater degrees of noise under intense time pressure. However, research on choice behavior has shown that when people are under time pressure, they often shift processing strategies to ensure reasonable accuracy (Ben Zur & Breznitz, 1981; Payne et al., 1988; Wright, 1974). We suspect that when making likelihood estimates in polychotomous cases, people placed under severe time pressure rely on heuristic comparison processes. In the raffle paradigm, we expected that time pressure would cause people to shift from trying to sum the evidence for alternative players (part of the normative process which is presumably resource intensive) to relying more on pairwise comparisons that are easier to execute (with the comparison between the focal and strongest alternative carrying inordinate weight). Hence, alternative-outcomes effects, which are products of the comparison heuristic, should increase under time pressure, even though people’s sensitivity to changes in the objective likelihood of the focal outcome might tend to be hampered by time pressure.

The evidence representation factor

The evidence on which likelihood judgments are based can be of various types. In the raffles we have discussed thus far, evidence is represented as a precise number. However, one can also make a likelihood

judgment based on similarity information (as in the famous Linda problem; Tversky & Kahneman, 1982), on memory for past events (see Windschitl, Young, & Jenson, 2002), or on other forms of information. We assume that the comparison heuristic can be applied in any type of likelihood judgment, irrespective of how a respondent went about judging the strength of the evidence for the possible outcomes (see Windschitl et al., 2002). In Experiments 1 and 2 of this paper, participants viewed raffles in which the evidence was represented by numbers (as in Figure 1a) or by stacks that were said to proportionally reflect quantities of tickets (as in Figure 1b). We expected that the alternative-outcomes effects observed in the graphical-representation condition would be as large or larger than those observed in the numerical-representation condition. Previous research on graph comprehension has identified ways in which graphical displays facilitate various forms of information processing or can be used to draw attention to specific trends, comparisons, or dimensions (e.g., Jarvenpaa, 1989, 1990; Schkade & Kleinmuntz, 1994; Shah, Mayer, & Hegarty, 1999; Stone et al., 2003; see review by Shah, Freedman, & Vekiri, in press). However, the research described in the graph-comprehension literature provides only clues for answering the question of whether a graphical display would influence the degree of heuristic processing that mediates a likelihood judgment. Our primary rationale for why we thought that the alternative-outcomes effect might be larger in the graphical-representation condition was that we assumed that participants would have difficulty summing the numbers of tickets in graphically represented raffles, because the precise number of tickets held by individuals was not explicitly stated and participants could not directly translate the height of a ticket stack into a specific number. If indeed participants had difficulty summing the tickets in a raffle, they might rely on the comparison heuristic, which does not require the respondent to sum evidence across alternatives. Also, because tabular representations of data tend to require greater serial and effortful processing than do graphical representations when information must be integrated (see Jarvenpaa & Dickson, 1988; Shah et al., in press), it seems reasonable to speculate that people might tend to approach graphical displays with a more effortless processing style—that is, a style in which they do not anticipate needing to think in a deliberative and precise manner (for a related argument, see Kleinmuntz & Schkade, 1993). This might lead to a rejection of the more deliberative, normative approach in favor of heuristic comparisons. Finally, graph-comprehension research by Simkin and Hastie (1987) suggests that bar-chart formats are particularly conducive to comparing among values (see also Jarvenpaa & Dickson, 1988). Although Simkin and Hastie were evaluating bar charts versus pie charts (not bar charts vs. tables), we nevertheless think it is notable that bar-chart displays were not particularly conducive to part-to-whole comparisons, comparisons which are essentially required for arriving at normative likelihood judgments about the raffles.

Individual differences in rational and experiential thinking

Although our main focus in this research concerned the influence of task conditions as potential moderators of processing approaches, we also tested in Experiment 1 the potential moderating role of individual differences in information processing by administering the Rational Experiential Inventory (REI; Pacini & Epstein, 1999). The REI measures a participant's self-proclaimed ability and engagement in both rational (rule-based) and experiential (intuitive) thinking. Agreeing with statements such as "I have a logical mind" and "I like to rely on my intuitive impressions" would reflect propensities for rational and intuitive thinking, respectively. To the extent that people generally prefer rule-based reasoning, they would presumably tend to approach the raffles in Experiment 1 from a rule-based perspective, which would yield normatively appropriate responses. People who prefer to rely on experiential or intuitive solutions might show an opposite tendency towards relying on heuristic solutions, such as the comparison heuristic. Hence, we predicted that the extent to which participants exhibited alternative-outcomes effects would be negatively correlated with their scores on the rational subscale of the REI and/or positively correlated with their scores on the experiential subscale.

EXPERIMENT 1

Overview

Experiment 1 used a methodology similar to that used by Windschitl and Young (2001) in which participants saw and provided likelihood estimates for a series of hypothetical raffles. The tickets in the raffles were manipulated in such a way that would allow us to test the extent to which participants' judgments of likelihood exhibited alternative-outcomes effects as well as normative effects linked to true changes in the likelihood of the focal outcome. When Windschitl and Young used this paradigm, all participants provided non-numeric "gut-level" estimates of their optimism about winning. Both alternative-outcomes effects and normative effects were detected, but no moderator variables were examined. In the present experiment, three potential moderator variables were tested: response format, time pressure, and evidence representation. Participants also completed the REI (Pacini & Epstein, 1999).

Method*Participants and design*

Students ($N = 168$) from a University of Iowa introductory psychology course were tested in groups ranging in size from 1 to 4. We employed a mixed design in which response format (numeric vs. non-numeric), time pressure (self-paced vs. rushed), and a counterbalancing factor were manipulated between subjects. Evidence representation (numerical vs. graphical), raffle type (baseline, flat, peaked, concentrated, and reduced-focal), and raffle set (1–6) were manipulated within subjects.

Procedure

All instructions and stimuli were presented to participants via computers, and all responses were made via mouse clicks. Participants first saw a sample raffle and instructions about how the raffles should be interpreted. Next, participants saw the scale they were going to use when providing their responses. Participants in the numeric-response condition saw a row of 21 adjacent buttons, each labeled with a percentage increasing in five percent intervals from 0% (coded as 0) to 100% (coded as 20). Those participants were instructed that we were interested in their careful assessment of the true likelihood of winning and were asked to click on the response that best reflected the true objective probability of winning. Participants in the non-numeric-response condition saw a thick line anchored with "extremely unlikely" and "extremely likely." This line was invisibly partitioned into 21 equally sized areas, allowing us to code participants' responses from 0 to 20. Those participants were instructed that we were interested in their general impression of how likely they were to win and were asked for their gut-level responses.

The basic instructions for the participants in the self-paced and rushed conditions were identical, except that the instructions for the rushed condition included the following:

However, you will have only 5 seconds to view each raffle, so you will have to respond very quickly. If you have not responded within 5 seconds, the raffle picture will disappear, and you should provide your response immediately. Although this task is not an easy one, please "hang in there" and do your very best throughout the entire experiment. There is a pause after every raffle, so feel free to use it if you need a rest.

After the instructions, all participants responded to one practice raffle and then proceeded to the critical raffles (described below). All participants responded to the same 30 raffles twice, once with tickets represented by numbers (the numerical-evidence condition; Figure 1a) and once with tickets represented as stacks (the graphical-evidence condition; Figure 1b). This sequence was counterbalanced across participants, and the order of the 30 raffles was independently randomized for each participant. Lastly, participants completed the REI (Pacini & Epstein, 1999).

Table 1. Numbers of tickets held by focal and alternative players in raffles from experiments 1 and 2

Raffle Set	Raffle Type	Focal Player	Alternative Players ^a					p(focal)
			1	2	3	4	5	
1	Baseline	17	8	5	5	5	4	0.386
	Flat	17	8	8	8	7	7	0.309
	Peaked	17	19	5	5	5	4	0.309
	Concentrated	17	38					0.309
	Reduced-Focal	12	8	5	5	5	4	0.308
2	Baseline	11	7	5	4	4	4	0.314
	Flat	11	7	7	7	6	6	0.250
	Peaked	11	16	5	4	4	4	0.250
	Concentrated	11	33					0.250
	Reduced-Focal	8	7	5	4	4	4	0.250
3	Baseline	9	11	6	5	5	5	0.220
	Flat	9	11	8	8	8	7	0.176
	Peaked	9	21	6	5	5	5	0.176
	Concentrated	9	42					0.176
	Reduced-Focal	7	11	6	5	5	5	0.179
4	Baseline	14	7	4	3	3	3	0.412
	Flat	14	7	5	5	5	4	0.350
	Peaked	14	13	4	3	3	3	0.350
	Concentrated	14	26					0.350
	Reduced-Focal	11	7	4	3	3	3	0.355
5	Baseline	15	18	6	5			0.341
	Flat	15	18	13	11			0.263
	Peaked	15	31	6	5			0.263
	Concentrated	15	42					0.263
	Reduced-Focal	10	18	6	5			0.256
6	Baseline	19	8	3	2			0.594
	Flat	19	9	8	8			0.432 ^b
	Peaked	19	20	3	2			0.432
	Concentrated	19	25					0.432
	Reduced-Focal	10	8	3	2			0.435

^aThis table shows the ticket numbers for alternative players in an order from highest (on the left) to lowest (on the right). However, the actual position-ordering of these numbers varied across sets (see Footnote 1 for more information).

^bDue to a programming error, some participants saw a 19-7-8-8 raffle instead of the intended 19-9-8-8 raffle. The difference between the two raffles is exceedingly small, did not appreciably influence any of the results, and is not discussed further.

Raffles

The 30 raffles that were used constituted 6 sets, with 5 raffles in each. Table 1 shows the raffles in this conceptually organized fashion. It should also be noted that the left-to-right ordering of the numbers within each raffle is depicted in a conceptually organized fashion that differs from the ordering seen by participants.¹

In designing each set, we first constructed a *baseline* raffle. The construction of the baseline raffles was somewhat arbitrary, except that we wanted the probability of winning to be kept low in order to avoid ceiling effects. We used this baseline raffle as a starting point for devising the four other raffles in the set—*flat*,

¹In Table 1, the numbers of tickets held by alternative players are ordered from highest (on the left) to lowest (on the right). However, the actual ordering of the alternative players varied across the sets but not within sets. For example, the actual ordering of the numbers for players in Set 1 was 17-5-4-5-8-5 for baseline, 17-8-7-7-8-8 for flat, 17-5-4-5-19-5 for peaked, 17-38 for concentrated, and 12-5-4-5-8-5 for reduced-focal. In previous research using these types of raffle displays, the ordering of the alternative players was manipulated and found to have no significant bearing on participants' optimism about winning (Windschitl & Young, 2001). The exact positioning of the alternative players can be obtained by writing the authors.

peaked, concentrated, and reduced-focal. For example, consider Raffle Set 1. In Set 1, the flat raffle was identical to the baseline, except that 11 tickets were added to the weakest alternative players, resulting in a flat distribution among alternative players (i.e., they all had a roughly similar numbers of tickets). The peaked raffle also was identical to the baseline, except that 11 tickets were added to the strongest alternative player, resulting in a peaked distribution among alternative players. The concentrated raffle was also identical to the baseline, except that 11 tickets were added and all the alternative tickets were held by (or concentrated into) one player. Finally, the reduced-focal raffle was identical to the baseline, except that the tickets held by the focal player were reduced such that the probability of winning the reduced-focal raffle was the same (or very nearly the same) as in the flat, peaked, and concentrated raffles. The raffles of the other sets were constructed by the same basic method. Hence, within a set, the non-baseline raffles all offered the same probability of winning, which was always less than the probability of winning the baseline raffle.

Comparisons between participants' responses to the baseline, flat, peaked, and concentrated raffles can be used as indexes of normative and alternative-outcomes effects. Namely, the difference in responses between baseline and flat raffles reflects the extent to which participants were sensitive to a change in the objective likelihood of winning. The difference in responses between the flat and peaked raffles (or between peaked and concentrated) reflects the extent to which participants were sensitive to manipulations of how evidence was distributed across alternative players. The reduced-focal raffles are less relevant than the others to the main hypotheses in this paper, and therefore we will not discuss the results for this raffle type in much detail.

Results and discussion

Based on preliminary analyses involving a full mixed-model ANOVA,² we determined that the raffle-set factor could be dropped from our analyses without influencing the central conclusions of the study. Hence, we collapsed across raffle sets to compute average responses within each type of raffle. The means for the 4 raffle types that are most relevant to our hypotheses are depicted in Figure 3—broken down by response format and time pressure. (As discussed shortly, the evidence representation factor did not systematically influence our key results, so we collapsed data across this factor in Figure 3.) Means and standard deviations for each raffle type, as a function of all three manipulated variables, are displayed in Appendix A.

In order to test the main hypotheses, we created three indexes by calculating difference scores for each participant between his/her mean responses to specific raffles:

1. Alternative-Outcomes-Effect Index 1 (AOE1) = $M_{\text{Flat Raffles}} - M_{\text{Peaked Raffles}}$
2. Alternative-Outcomes-Effect Index 2 (AOE2) = $M_{\text{Peaked Raffles}} - M_{\text{Concentrated Raffles}}$
3. Normative-Change Effect Index (NCE) = $M_{\text{Baseline Raffles}} - M_{\text{Flat Raffles}}$

For brevity, we do not report results for another possible AOE index (based on the difference between the flat and concentrated raffles) because it produced results similar to the other AOE indexes. Also, although we could include more NCE indexes (e.g., based on the difference between baseline and peaked raffles) we omitted such indexes because they conflate a participant's sensitivity to a change in objective probability with his/her sensitivity to a change in the distribution of evidence.

As a preliminary comment, we note that one-sample *t*-tests (collapsed across all factors) show that the values in the AOE1, AOE2, and NCE Indexes were significantly different from zero, with all *p* values less than 0.001. In large measure then, participants were sensitive to changes in the distribution of evidence

²These preliminary analyses are less relevant to our main hypotheses than are the analyses reported in the body of the paper, yet we will briefly summarize their findings. In the full mixed-model ANOVA (on raw likelihood-judgment data) that included all possible factors, the main effects for raffle set and raffle type were, not surprisingly, quite robust ($p < 0.001$). The time-pressure main effect was not significant but the evidence-representation main effect was significant ($p < 0.05$), with participants giving slightly higher estimates in the numeric-representation condition (see Appendix B). The response-format main effect was significant ($p < 0.001$), but this effect is not very meaningful given that the numeric and non-numeric scales are not directly comparable (see Windschitl & Wells, 1996).

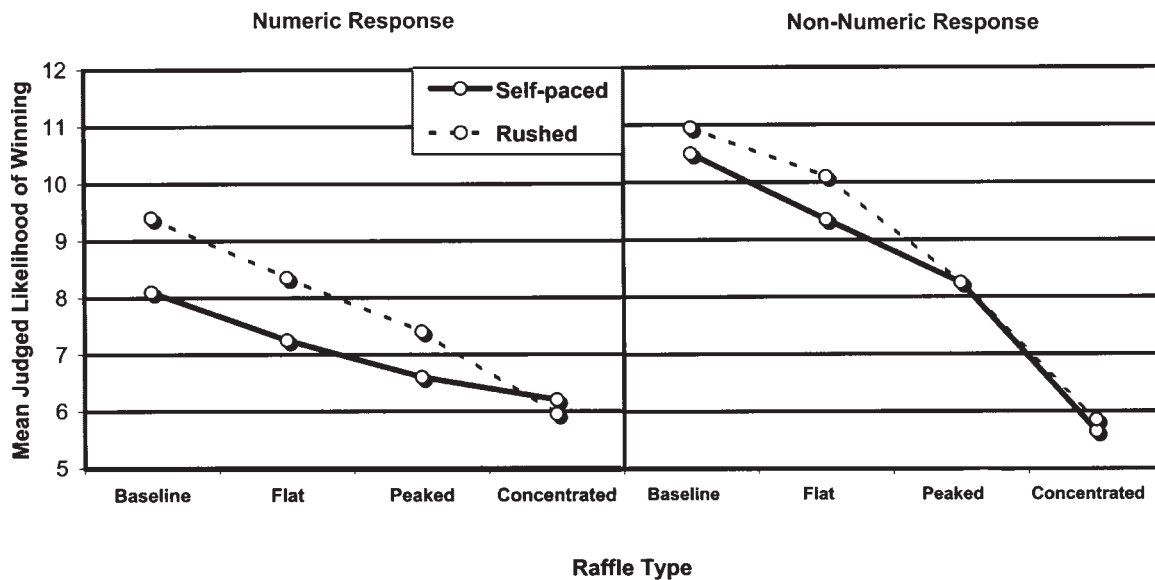


Figure 3. Mean judged likelihood of winning (scored from 0–20) as a function of response type, time pressure, and raffle type in Experiment 1. The *objective* probability of winning is higher in baseline raffles than in the other raffles. The flat, peaked, and concentrated raffles differ only in how tickets are distributed among alternative players

across alternative players (changes that do not influence a participant's chances of winning) and sensitive to the addition of tickets to alternative players (additions that do influence a participant's chance of winning).

The key analyses focus on how the three indexes, particularly the AOE indexes, were influenced by the response-format, time-pressure, and evidence-representation manipulations. The values for the AOE1 and AOE2 Indexes were submitted to separate ANOVAs with evidence representation as a within-subjects factor and response format and time pressure as between-subjects factors. As predicted, the main effect for the response-format factor was significant in both analyses, $F(1, 164) = 8.37, p < 0.01$ and $F(1, 164) = 19.33, p < 0.001$, respectively. Participants responding on non-numeric scales showed much larger alternative-outcomes effects ($M_{AOE1} = 1.42, SD = 1.59; M_{AOE2} = 2.53, SD = 2.76$) than did participants responding on numeric scales ($M_{AOE1} = 0.80, SD = 1.22; M_{AOE2} = 0.91, SD = 1.99$). Consistent with our prediction regarding time pressure, the main effect for time pressure was significant in the analysis of the AOE1 Index, $F(1, 164) = 7.19, p < 0.01$. That is, the alternative-outcomes effect represented by the AOE1 Index (contrasting responses to flat versus peaked raffles) was larger for participants under time pressure ($M_{AOE1} = 1.39, SD = 1.50$) than for those who were self-paced ($M_{AOE1} = 0.82, SD = 1.35$). Data from the AOE2 Index (peaked versus concentrated raffles) show a similar but non-significant trend, $F(1, 164) = 1.40, p = 0.24, (M_{AOE2} = 1.93, SD = 2.57; M_{AOE2} = 1.50, SD = 2.49, respectively)$. We had also predicted that alternative-outcomes effects in the graphical-representation condition would be as large or larger than those in the numerical-representation condition. The main effect for this factor was not significant in the analysis of the AOE1 Index, $F(1, 164) = 0.38, p = 0.54$. However, it was significant in the analysis of the AOE2 Index, $F(1, 164) = 7.58, p < 0.01$, with the alternative-outcomes effect being somewhat larger in the numerical ($M_{AOE2} = 1.98, SD = 2.93$) rather than graphical condition ($M_{AOE2} = 1.45, SD = 2.71$). Finally, of the 8 possible interactions across these two ANOVAs, none were statistically significant.

To summarize the findings from ANOVAs on alternative-outcomes indexes, participants displayed larger alternative-outcomes effects when responding on a non-numeric scale and when answering under time pressure. Although there was one analysis suggesting that alternative-outcomes effects might be larger when

evidence is represented numerically rather than graphically, this finding was not replicated in Experiment 2, and therefore we do not give it much credence.

Although the alternative-outcomes effect indexes showed sensitivity to the response-format and time-pressure factors, the normative-change index was not influenced by these factors nor the evidence-representation factor. Specifically, the NCE Index yielded no significant main effects or interactions (all $p > 0.05$).

Response latencies

As would be expected, response latencies were considerably shorter and less variable in the time-pressure condition ($M = 2.92$ s, $SD = 0.56$ s) than in the self-paced condition ($M = 5.92$ s, $SD = 3.30$ s). For the sake of brevity, we report only our analyses for the self-paced condition; the results from the time-pressured condition were generally similar, but less extreme due to the increased uniformity created by the response deadline. Figure 4 depicts the mean response latencies (untransformed) from the self-paced condition for each of the raffle types that are most relevant to our main hypotheses.

Log-transformed response latencies from the self-paced condition were submitted to a mixed-model ANOVA with raffle set, raffle type, and evidence representation as within-subjects factors; response format was the only between-subjects factor. The raffle-set and raffle-type main effects were significant, $F(5, 410) = 4.50$, $p < 0.001$, and $F(4, 328) = 49.28$, $p < 0.001$, respectively. These findings reflect the fact that participants responded faster to raffles that had fewer alternative players (i.e., raffles in Sets 5 and 6, and all concentrated raffles). More important, however, was the fact that the evidence-representation and the response-format main effects were both significant, $F(1, 82) = 20.57$, $p < 0.001$, and $F(1, 82) = 34.96$, $p < 0.001$, respectively. Participants were quicker to respond on non-numeric ($M = 4.26$ s, $SD = 1.79$ s) than on numeric scales ($M = 7.59$ s, $SD = 3.63$ s), and they were quicker when the evidence was represented graphically ($M = 5.16$ s, $SD = 3.24$ s) rather than numerically ($M = 6.51$ s, $SD = 3.86$ s).

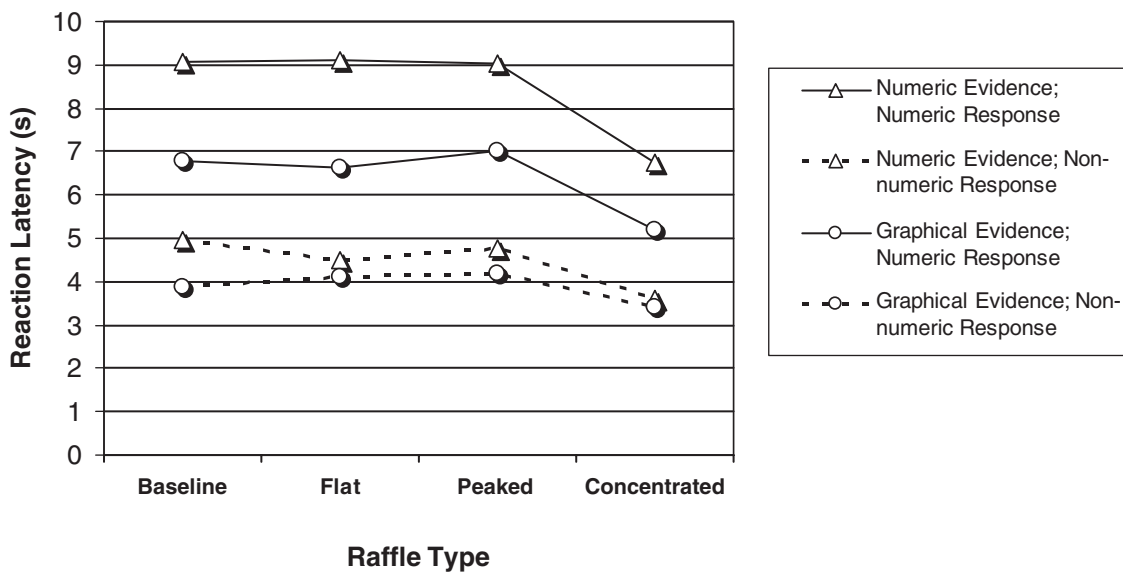


Figure 4. Response latencies as a function of evidence representation, response type, and raffle type from the self-paced condition in Experiment 1

The fact that participants giving numeric responses were relatively slow, whereas participants giving non-numeric responses were relatively fast, is consistent with the claim by Windschitl and Wells (1996) that participants who are asked to give numeric responses, relative to those asked to give non-numeric ones, engage in more deliberative processing of information. However, this difference in response time might also be due to the different instructions given to the two groups of participants—an issue that was resolved in Experiment 2.

The fact that participants were quicker to respond when ticket quantities were presented graphically rather than numerically is generally consistent with two possibilities we mentioned earlier: 1) that the imprecise representations of ticket quantities in the graphical condition would make a normative approach difficult and cause people to employ a heuristic approach, and 2) that because people typically expect graphs to be easily interpreted (see Kleinmuntz & Schkade, 1993), they would approach graphical displays with a relatively effortless and heuristic processing style. However, the likelihood judgment data do not support either of these possibilities; graphical representations of evidence, relative to numerical representations, were not more likely to prompt alternative-outcomes effects. Hence, it appears that people can execute likelihood-estimation processes more efficiently when evidence is represented graphically rather than numerically, regardless of whether these processes are largely heuristic or rule-based. Experiment 2 provides an opportunity to see if the findings supporting this unanticipated conclusion are replicated.

Finally, we also examined the relation between response latencies and the AOE1, AOE2, and NCE Indexes for participants in the self-paced condition. The correlation between response latencies and the NCE Index was positive but nonsignificant, $r = 0.10$, $p > 0.10$. However, the correlations between response latencies and the alternative-outcomes-effect indexes were significantly negative, $r_{\text{AOE1}} = -0.31$, $p < 0.01$; $r_{\text{AOE2}} = -0.46$, $p < 0.001$. These negative correlations indicate that people who responded quickly tended to use the comparison heuristic.

Rational-experiential inventory

Although our main hypotheses focused on how specific task conditions influence processes underlying likelihood judgments, data from the Rational Experiential Inventory (REI) also allow us to examine potential individual differences in these processes. Consistent with our prediction that participants who prefer to engage in rule-based thinking would be less sensitive to alternative-outcomes effects, scores on the rational subscale of the REI were correlated negatively with both the AOE1 and AOE2 index, $r(167) = -0.22$, $p < 0.01$, and $r(167) = -0.18$, $p < 0.02$, respectively. Scores on the rational subscale were also (marginally) positively correlated with the NCE index, $r(167) = 0.13$, $p = 0.08$. Preference to engage in experiential thinking, however, was not systematically related to any of our indexes (all $p > 0.10$). Although the results of the individual difference measures are somewhat mixed, they do provide support for the notion that when people are not inclined to take a normative/rule-based approach to solving a likelihood-judgment problem, this does not necessarily lead to random error in the judgment process, but instead leads to an increased reliance on a systematic yet heuristic process that yields alternative-outcomes effects.

EXPERIMENT 2

The findings from Experiment 1 left some issues unresolved. For example, although the results clearly showed that alternative-outcomes effects can be observed regardless of whether evidence is represented numerically or graphically, the results were mixed as to whether numerical and graphical representations of evidence prompt differing amounts of alternative-outcomes effects. To address this and other issues more definitively, we conducted a replication of Experiment 1 that included largely similar methods, factors, and materials.

We again manipulated the evidence representation factor in order to assess whether graphical representation of evidence prompts the same or different levels of alternative-outcomes effects. To allow for the cleanest possible manipulation of this factor, we used a between-subjects manipulation, rather than the within-subjects manipulation that was used in Experiment 1.

In Experiment 1, time pressure appeared to increase the use of the comparison heuristic, as evidenced by a significant effect on the AOE1 Index. However, although the same type of effect was observed for the AOE2 Index, this effect was not significant. Hence, we wished to explore the role of time pressure again. In Experiment 2, we made the time pressure manipulation a bit more severe—giving participants only 3 seconds to provide their responses.

The response-format main effects in Experiment 1 were quite strong. Participants exhibited larger alternative-outcomes effects on the non-numeric than on numeric scales, and they took longer to decide on a numeric response than a non-numeric one. Although both of these findings are consistent with the idea that numeric measures are more likely than non-numeric ones to prompt deliberative and rule-based thinking (Windschitl & Wells, 1996), the results might alternatively be explained by the difference in the instructions that accompanied the two types of response scales. Participants in the non-numeric condition, but not the numeric condition, were encouraged to indicate their gut-level impressions when providing a response (similar to instructions used by Windschitl & Young, 2001). To assess whether the response scales themselves produce differential processing of information, we gave all participants in Experiment 2 the same instructions about the response scales. In fact, we went to some lengths to encourage all participants to remain highly motivated and focused throughout the experiment while trying to provide the most objectively correct or appropriate answer they could give for each raffle.

Method

Participants, design, and raffles

Students ($N = 301$) from a University of Iowa introductory psychology course were tested in groups ranging in size from 1 to 4. The design was the same as that used in Experiment 1, except the evidence-representation factor (numeric vs. graphical) was manipulated between subjects. The raffles were identical to those in Experiment 1.

Procedures

The procedures were identical to those used in Experiment 1 with some exceptions. First, 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 . . . Please try to give the most correct or appropriate answer that you can for each raffle, even the last one.

Before beginning the experiment, participants were given an opportunity to inform the experimenter if they thought they would be unable to make a full effort on all the raffles (none took this opportunity). To reinforce the notion that we wanted their careful assessments, we also offered people a pencil and sheet of paper and told them “You might find this paper helpful at various points in this experiment. Feel very free to use the paper for any purposes you deem necessary.” A second difference from Experiment 1 involved the response scale instructions. Although the same numeric and non-numeric scales were used in Experiment 2, the instructions solicited participants’ assessments of the true likelihood of winning regardless of the response format. A third difference was that the response deadline in the time-pressure condition was changed from 5 to 3 seconds. A fourth difference was that, because of time constraints within our experimental sessions, participants did not complete the REI.

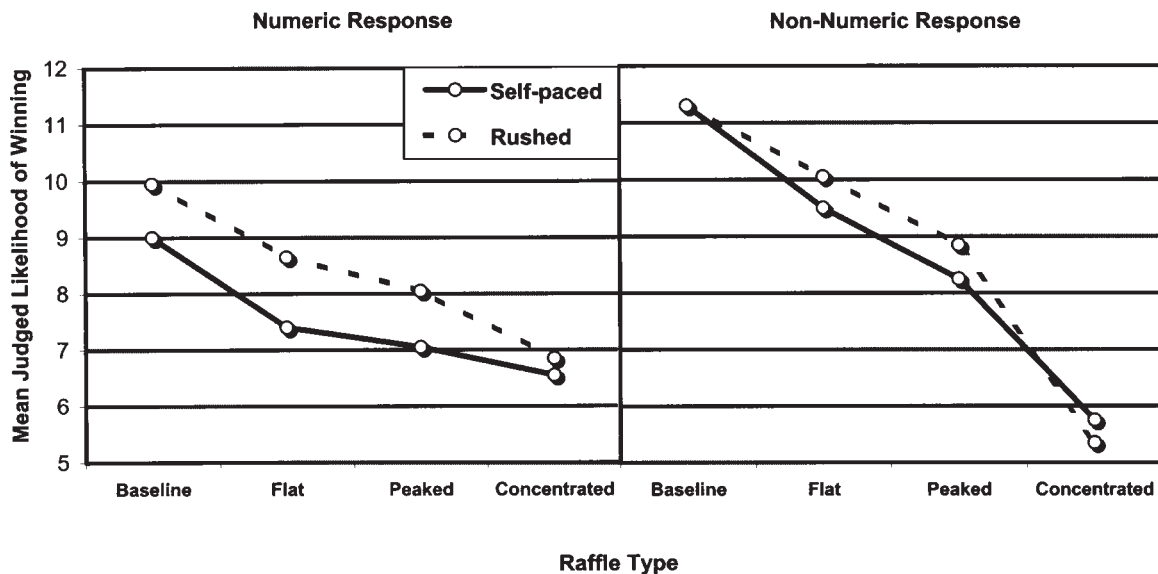


Figure 5. Mean judged likelihood of winning (scored from 0–20) as a function of response type, time pressure, and raffle type in Experiment 2. The *objective* probability of winning is higher in baseline raffles than in the other raffles. The flat, peaked, and concentrated raffles differ only in how tickets are distributed among alternative players

Results and discussion

Results of preliminary analyses involving a full mixed-model ANOVA closely resembled those of Experiment 1.³ Hence, we again collapsed across raffles sets to compute average responses within each type of raffle. The means for the baseline, flat, peaked, and concentrated raffles are depicted in Figure 5. Means and standard deviations for all raffles are in Appendix B. We will again report the analyses for the three indexes that we created: AOE1, AOE2, and NCE. As in Experiment 1, preliminary one-sample *t*-tests (collapsed across all factors) showed that the mean values of these indexes were all significantly different from zero ($p < 0.001$).

The values for the AOE1 and AOE2 Indexes were submitted to separate ANOVAs with response format, time pressure, and evidence representation as between-subjects factors. The main effect for the response-format factor was again significant in both analyses, $F(1, 294) = 11.82$, $p < 0.001$ and $F(1, 294) = 33.26$, $p < 0.001$, respectively. Participants responding on non-numeric scales showed much larger alternative-outcomes effects ($M_{AOE1} = 1.24$, $SD = 2.15$; $M_{AOE2} = 2.99$, $SD = 3.63$) than did participants responding on numeric scales ($M_{AOE1} = 0.48$, $SD = 1.63$; $M_{AOE2} = 0.88$, $SD = 2.67$). Unlike in Experiment 1, the main effect for time pressure was not significant for the AOE1 Index, $F(1, 294) = 0.162$, $p = 0.70$. However, the same main effect was significant for the AOE2 Index, $F(1, 294) = 5.59$, $p < 0.05$, with participants in the time-pressure condition ($M = 2.37$, $SD = 3.45$) showing larger alternative-outcome effects than those in the self-paced condition ($M = 1.51$, $SD = 3.21$). Finally, both of the evidence-representation main effects were not significant, both F 's < 1 . None of the eight interactions within these two ANOVAs were significant.

The results from the ANOVA on the NCE Index revealed a somewhat different pattern. The main effect for the response-format factor was not significant, $F < 1$. However, unlike in Experiment 1, the time-pressure

³In the preliminary analyses, the mixed-model ANOVA (on raw data) that included all possible factors revealed significant main effects for raffle type, raffle set, and response format (all p 's < 0.05) but nonsignificant effects for evidence representation and time pressure.

main effect was significant, $F(1, 294) = 5.79, p < 0.05$. Participants under time pressure were *less* sensitive to the change in objective probability ($M = 1.26, SD = 1.64$) than were participants who were self paced ($M = 1.72, SD = 1.60$). The evidence-representation main effect was not significant, $F < 1$. None of the interactions were significant.

With regards to the main hypotheses under investigation, the results from Experiment 2 support the same general conclusions as do the results from Experiment 1. The significant effects for response format on the AOE Indexes provide clear support for the idea that people tend to rely on the comparison heuristic to a greater degree when they are providing non-numeric rather than numeric likelihood judgments. These effects were quite robust even though participants were given the exact same instructions about how to use the two scales. Hence, it appears that the numeric features of the response scale do indeed elicit a different set of judgment processes than do the features of a non-numeric scale (Windschitl & Wells, 1996). Placing participants under time pressure increased their tendency to rely on the comparison heuristic (as evidenced by the effect on the AOE2 index), but it also significantly decreased their sensitivity to a change in the objective probability of the focal outcome (as assessed by the NCE Index). Finally, it appears that the way in which evidence was represented did not have much impact on the key likelihood judgment processes under investigation; in Experiment 2, the alternative-outcomes effects were just as large (and no larger) when evidence was presented graphically rather than numerically.

Response latencies

Response latencies were again considerably shorter and less variable in the time-pressure condition ($M = 2.90$ s, $SD = 2.75$ s) than in the self-paced condition ($M = 11.01$ s, $SD = 8.10$ s). As in Experiment 1, an ANOVA on log-transformed latency data from the self-paced condition revealed significant raffle-set and raffle-type main effects, $F(5, 735) = 14.26, p < 0.001$, and $F(4, 588) = 69.19, p < 0.001$. More importantly, the evidence-representation and the response-format main effects were again significant, $F(1, 147) = 19.79, p < 0.001$, and $F(1, 147) = 20.38, p < 0.001$. Participants were quicker to respond on non-numeric ($M = 8.36$ s, $SD = 5.92$) than on numeric scales ($M = 13.70$ s, $SD = 9.11$), and they were quicker when evidence was represented graphically ($M = 8.38$ s, $SD = 5.73$) rather than numerically ($M = 13.62$ s, $SD = 9.22$).

Figure 6 depicts the response latencies from the self-paced condition, for each of the raffle types that are most relevant to our main hypotheses. This pattern is nearly identical to that observed for Experiment 1. Unlike in Experiment 1, the quicker response latencies for non-numeric rather than numeric responses cannot be attributed to differences in the instruction that accompanied the two types of scales. Hence, the response latency data from Experiment 2 provide clear support for the claim by Windschitl and Wells (1996) that participants who are asked to give numeric rather than nonnumeric responses engage in more deliberative processing of information.

The data from Experiment 2 support our conclusions from Experiment 1 about the relation between processing approaches and response latencies for graphical and numerical raffles. The fact that alternative-outcomes effects were no larger in the graphical than numerical conditions precludes an account suggesting that graphical representations of evidence prompted an increase in the use of the comparison heuristic, which thereby lowered response latencies. Instead, it appears that, regardless of the processes that are used to generate a likelihood estimate, people can execute these processes more efficiently when evidence is represented graphically rather than numerically.

Finally, we again examined the relations between response latencies and the AOE1, AOE2, and NCE Indexes (within the self-paced condition). As in Experiment 1, the correlation involving the NCE Index was positive but non-significant, $r = 0.08, p > 0.10$. The correlations involving the AOE1 and AOE2 Indexes were both negative and significant, $r_{AOE1} = -0.28, p < 0.001$; $r_{AOE2} = -0.36, p < 0.001$. It again appears that people who responded quickly tended to use the comparison heuristic.

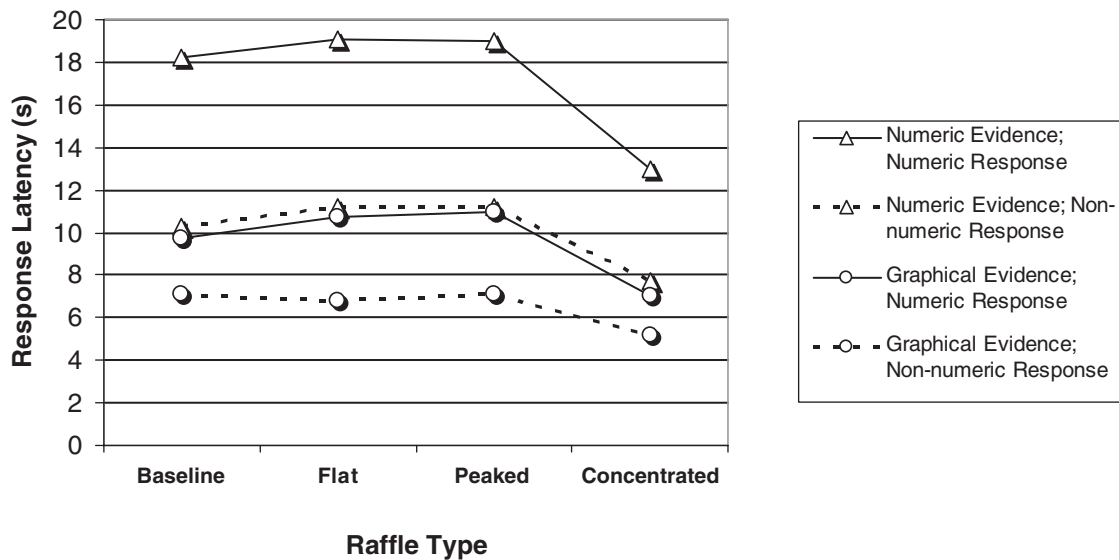


Figure 6. Response latencies as a function of evidence representation, response type, and raffle type from the self-paced condition in Experiment 2

CORRELATIONAL ANALYSES FOR EXPERIMENTS 1 AND 2

Thus far, we have presented results from an ANOVA framework, which tested the mean differences between specific types of raffles. However, the data from the two experiments can also be analyzed from a correlational approach. Instead of defining an NCE Index as the extent to which participants' responses to baseline raffles differed from their responses to flat raffles, we can define a *Normative Index* as the extent to which participants' responses across the raffles correlated with what a normative model would predict. Also, instead of defining an AOE Index as the extent to which participants' responses to flat raffles differed from their responses to peaked raffles, we can define a *Comparison-Heuristic Index* as the extent to which participants' responses to all raffles correlated with what a comparison-heuristic model would predict.

Hence, for each raffle, we calculated predictions for a normative model and a comparison-heuristic model. Assume that F represents the strength of the evidence for the focal outcome and that A_j represents the strength of the evidence for the j th strongest alternative. The normative model was calculated as: $F / (F + A_1 + A_2 + A_3 + A_4 + A_5)$. Although the comparison-heuristic account assumes multiple pairwise comparisons occur (with the focal-vs.-strongest-alternative playing a dominating role) our model simplified this account as a comparison between only the focal and the strongest alternative outcome. That is, the comparison-heuristic model was calculated as: $F / (F + A_1)$.⁴ Next, for each participant we computed 2 correlations reflecting how his/her likelihood judgments across the raffles related to 1) the predicted values from the normative model and 2) the predicted values from the comparison-heuristic model. The resulting correlation coefficients constitute a *Normative Index* and *Comparison-Heuristic Index*, respectively. After transforming these correlation coefficients (using Fisher's r to z transformations), we submitted them as data points to ANOVAs (one for each experiment) with response format, time pressure, evidence representation, and index (normative or comparison heuristic) as factors.

⁴For the raffles that were used in these experiments, the correlation between the values for the normative and comparison-heuristic model was 0.64.

Table 2. Means for the normative index and comparison-heuristic index collapsed across experiments 1 and 2

	Normative Index		Comparison-Heuristic Index	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Numeric Condition				
Self-Paced	0.67	0.22	0.46	0.23
Time-Pressured	0.58	0.19	0.51	0.26
Non-numeric Condition				
Self-Paced	0.61	0.22	0.60	0.27
Time-Pressured	0.56	0.19	0.60	0.27

Note: A value on the Normative Index reflects the extent to which a participant's responses across the 30 raffles correlated with the objective probability of winning those raffles. A value on the Comparison-Heuristic Index reflects the extent to which a participant's responses correlated with output from a comparison-heuristic model (which estimates likelihood by comparing the number of tickets held by the participant to the number of tickets held by the strongest alternative player). The mean values for these indexes are shown in this table.

The ANOVA for Experiment 1 revealed no significant main effects other than a significant effect for index, $F(1, 164) = 9.85$, $p < 0.01$, which simply reflects that—across the entire design—participants' likelihood judgments correlated with the normative model more than they correlated with the comparison-heuristic model. More important is that the index factor interacted significantly with the response-format factor, $F(1, 164) = 17.05$, $p < 0.001$, and with the time-pressure factor, $F(1, 164) = 7.12$, $p < 0.01$. The pattern of significant findings was nearly identical for the same analysis on data from Experiment 2 (an additional main effect for scale was found). Given the similarity in the results for Experiments 1 and 2, we combined the data from those experiments in order to create Table 2, which depicts how the values of the Normative Index and Comparison-Heuristic Index differ as a function of response format and time pressure.

These results provide broad confirmation of the findings from the analyses already described. The Index \times Response-Format interaction suggests that a numeric response scale has different effects on people's tendencies to use a normative and heuristic approach. Whereas the numeric responding promotes the use of a normative approach, it seems to reduce the use of a comparison-heuristic approach. The Index \times Time Pressure interaction reveals that there is a similar, although somewhat less pronounced, influence of time pressure.

GENERAL DISCUSSION

We set out to investigate how three potential moderator variables—response format, time pressure, and evidence representation—influence likelihood judgments regarding polychotomous cases. More specifically, rather than focusing on whether these variables had a main effect influence on likelihood judgments (making judgments generally higher or lower), we focused on how these three variables influenced three types of mediating approaches to likelihood judgment: 1) a normative approach that involves the application of learned rules of probability, 2) a comparison-heuristic approach in which likelihood judgment is heavily influenced by a pairwise comparison between the evidence for the focal and strongest alternative outcome, and 3) a random approach. There was clear evidence from our experiments that two of the three variables can have a moderating role on the relative influences of these approaches.

Consistent with previous work, the non-numeric response scale led to larger alternative-outcomes effects than did the numeric response scale (see Windschitl & Wells, 1998). Non-numeric responses were also made substantially faster than were numeric ones. These findings support the argument that when people are asked to give a non-numeric likelihood response (as opposed to a numeric one), they utilize a quicker and less deliberative judgment process in which a pairwise comparison between the focal and the strongest alternative outcome plays a critical role (i.e., the comparison heuristic).

Time pressure had a related influence. Overall, people were quite good at making likelihood judgments under time pressure. Even under the heavy time pressure in Experiment 2, the extent to which participants' numeric responses correlated with a normative model remained relatively high ($r = 0.63$). Nevertheless, time pressure did hurt the extent to which people's likelihood judgments were sensitive to changes in objective likelihood (Exp 2), presumably because it is difficult to accurately sum the evidence for alternative players (which is part of the normative process) under intense time pressure. Time pressure did not hurt the extent to which people's likelihood judgments were sensitive to changes in the distribution of evidence across alternatives. In fact, the results from the experiment suggest that participants increased their reliance on the comparison heuristic when under time pressure. This finding fits with previous work from outside the likelihood judgment domain showing that people often shift processing strategies to ensure reasonable accuracy under time pressure or distraction (Ben Zur & Breznitz, 1981; Payne, Bettman, & Johnson, 1988; Wright, 1974).

Overall, the use of the comparison heuristic was equally robust when people saw graphical or numerical representations of ticket quantities. Another key finding was that people were quicker at making judgments from the graphical displays than from the numerical displays. Recall that we had speculated that graphical displays might prompt low-effort processing (more heuristic processing) and that respondents would have a difficult time summing the tickets in the raffle, which is an operation required by the normative approach. However, the fact that respondents were quicker when faced with graphical rather than numeric displays—yet there was no difference in heuristic processing—suggests that the graphical display format was easier to interpret and use, regardless of whether a respondent employed a normative or heuristic approach to judging likelihood. Perhaps even a normative approach was made easier by the graphical displays, by allowing respondents to visually simulate the length of all stacks combined. This visual simulation would achieve the same goal as performing mental arithmetic to sum the overall number of tickets in the raffle. Graphical displays that do not allow people to easily visually simulate an additive representation of evidence might produce findings different from those in Experiments 1 and 2.

Related findings regarding the dud-alternative effect

Recent research on a phenomenon related to the alternative-outcomes effect has produced findings that parallel some of the findings from Experiments 1 and 2 (Windschitl & Chambers, 2004). This related phenomenon, called the dud-alternative effect, is defined as an increase in the judged likelihood of a focal outcome when very weak alternatives (“duds”) are added to the list of alternative outcomes. In one experiment, participants' non-numeric likelihood estimates about winning a raffle were more optimistic when they held 20 tickets and other players held 16, 2, and 1 tickets than when they held 20 tickets and the only other player held 16. This effect, like the alternative-outcomes effect, appears to be a product of a pairwise comparison approach to judging likelihood. Even though the pairwise comparison between the focal and strongest alternative outcome has a substantially disproportionate influence on a likelihood judgment (which causes alternative-outcomes effects), the pairwise comparisons between the focal and weaker alternative outcomes also have some influence—an influence that can result in dud-alternative effects. For example, in the dud-present raffles mentioned above, the number of tickets held by the respondent seems like much more than those held by either of the two dud players (those holding 2 and 1 tickets). These are two comparisons that are quite favorable to the focal outcome. However, these two favorable comparisons are not relevant (and not even available) when thinking about the dud-absent raffle. Hence, in terms of pairwise comparisons, the focal outcome fares better in the case of the dud-present raffle than in the case of the dud-absent raffle (see Windschitl & Chambers, 2004).

A key parallelism in the findings from the present research and the research on the dud-alternative effect is that dud-alternative effects, like alternative-outcomes effects, were found to be sensitive to both response format and time pressure (graphical representations were not tested). The dud-alternative effect was robust when detected by a non-numeric likelihood measure but was reversed when participants were asked to give

numeric likelihood responses. The effect was also significantly enhanced by the same type of time pressure applied in the second experiment of this paper.

Implications for theories of likelihood judgment

Taken together, the findings from the present research and that on the dud-alternative effect make a strong case that theories of likelihood judgment must attempt to consider how moderating variables such as response format and time pressure influence judgment processes. Likelihood judgments do not simply degrade in the direction of randomness when respondents are confronted with a non-numeric scale or with time pressure. Instead, respondents begin to rely more on a nonnormative but still systematic pairwise comparison approach to judging likelihood.

Existing theories of likelihood judgment that can be applied to polychotomous cases do not account for alternative-outcomes effects or dud-alternatives effects, and they do not specify how likelihood judgment processes might change as a function of important task conditions. A leading theory, support theory, can explain a variety of phenomena through its notion of subadditive support (see Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994). However, it is not specific (beyond types of subadditivity) as to how evidence/support assessments for individual alternative hypotheses (all of them in the relevant set) are treated or integrated. Hence, support theory in its present form is best considered a uni-process model. The present work suggests that a contingent-process model might prove beneficial for understanding and predicting people's likelihood judgments in polychotomous cases. Whether this contingent-process model should build upon support theory or should be developed as distinct from support theory is an open question.

Parallels/nonparallels to contingent decision making

In reviewing research on constructive decision making, Payne, Bettman, and Johnson (1992) described both a cost/benefit framework and a perceptual framework for contingent processing. Accounts falling within the cost/benefit framework assume that the selection of a processing strategy (e.g., lexicographic) depends on a weighing of the costs and benefits of possible strategies given the task conditions. Accounts falling within the perceptual framework assume that the use of strategies is influenced by perceptual or representational features of the task and environment. Whether this distinction will be a useful one for understanding contingent processes in likelihood judgments about polychotomous cases is difficult to determine. The fact that time pressure influenced the utilization of the normative and comparison-heuristic approaches seems to fall under the cost/benefit framework. However, the fact that the response format influenced the relative use of the normative and heuristic approaches is difficult to classify into one framework. On one hand, the response-format effect seems to fall under the perceptual framework if we assume that a numeric response format, but not the non-numeric format, somehow primed formal rules for determining likelihood. On the other hand, the response-format effect seems to fall under the cost/benefit framework if we assume that the non-numeric response format lowered the perceived costs of being somewhat inaccurate, if participants knew that the possible response values could not be directly mapped to numbers and, therefore, could not be deemed correct or incorrect. Hence, it is perhaps best to remain agnostic as to whether the cost/benefit versus perceptual framework distinction that was useful for describing contingent decision making will be equally useful for organizing future work on contingent processes in likelihood judgment.

ACKNOWLEDGEMENT

This work was supported by Grant SES 99-11245 to Paul D. Windschitl from the National Science Foundation.

APPENDIX A

Mean judged likelihood of winning as a function of response type, evidence representation, time pressure, and raffle type in Experiment 1. Standard deviations appear within parentheses. Responses on both the numeric and nonnumeric scales were scored from 0 (0%, extremely unlikely) to 20 (100%, extremely likely).

Raffle Type	Numeric Response				Non-numeric Response			
	Numeric Evidence		Graphical Evidence		Numeric Evidence		Graphical Evidence	
	Self-paced	Rushed	Self-paced	Rushed	Self-paced	Rushed	Self-paced	Rushed
Baseline	8.53 (2.78)	9.58 (3.15)	8.07 (2.47)	9.25 (3.05)	10.73 (3.48)	11.19 (3.03)	10.30 (2.96)	10.65 (3.40)
Flat	7.30 (3.00)	8.36 (3.25)	7.21 (2.48)	8.29 (3.19)	9.39 (3.98)	10.30 (3.80)	9.07 (3.55)	9.94 (3.20)
Peaked	6.56 (2.16)	7.31 (2.65)	6.65 (2.23)	7.46 (2.35)	8.63 (3.14)	8.62 (2.90)	7.85 (2.92)	7.93 (2.56)
Concentrated	6.01 (1.47)	5.81 (1.43)	6.42 (1.11)	6.10 (1.59)	5.66 (1.95)	5.71 (2.17)	5.61 (2.56)	5.96 (2.14)
Reduced-Focal	6.98 (2.46)	8.01 (3.01)	6.72 (2.57)	7.34 (2.58)	9.20 (3.43)	9.51 (2.87)	7.92 (2.47)	8.71 (2.91)

APPENDIX B

Mean judged likelihood of winning as a function of response type, evidence representation, time pressure, and raffle type in Experiment 2. Standard deviations appear within parentheses. Responses on both the numeric and nonnumeric scales were scored from 0 (0%, extremely unlikely) to 20 (100%, extremely likely).

Raffle Type	Numeric Response				Non-numeric Response			
	Numeric Evidence		Graphical Evidence		Numeric Evidence		Graphical Evidence	
	Self-paced	Rushed	Self-paced	Rushed	Self-paced	Rushed	Self-paced	Rushed
Baseline	9.02 (2.23)	9.72 (2.65)	9.02 (2.65)	10.16 (3.03)	11.06 (3.03)	11.11 (3.66)	11.54 (4.00)	11.47 (3.26)
Flat	7.60 (2.81)	8.47 (3.05)	7.23 (2.41)	8.85 (3.40)	9.39 (3.56)	9.76 (3.80)	9.58 (3.91)	10.35 (3.60)
Peaked	7.27 (2.55)	7.60 (2.36)	6.84 (2.04)	8.51 (3.01)	8.15 (2.71)	8.90 (3.10)	8.28 (3.13)	8.80 (2.82)
Concentrated	6.20 (1.94)	6.36 (1.59)	6.85 (1.36)	7.30 (2.30)	5.90 (1.71)	5.30 (3.29)	5.60 (2.12)	5.40 (2.30)
Reduced-Focal	7.67 (2.65)	8.72 (2.91)	7.43 (2.37)	8.20 (3.14)	9.06 (3.05)	9.09 (3.31)	9.06 (3.17)	8.84 (2.52)

REFERENCES

- Ben Zur, H., & Breznitz, S. J. (1981). The effects of time pressure on risky choice behavior. *Acta Psychologica*, *47*, 89–104.
- Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: processes of judgment and choice. *Annual Review of Psychology*, *32*, 53–88.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: beyond “Heuristics and biases.” *European Review of Social Psychology*, *2*, 83–115.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: a reply to Kahneman and Tversky (1996). *Psychological Review*, *103*, 592–596.
- GINOSSAR, Z., & TROPE, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology*, *52*, 464–474.
- Gonzalez, M., & Frenck-Mestre, C. (1993). Determinants of numerical versus verbal probabilities. *Acta Psychologica*, *83*, 33–51.
- Jarvenpaa, S. L. (1989). The effect of task demands and graphical format on information processing strategies. *Management Science*, *35*, 285–303.
- Jarvenpaa, S. L. (1990). Graphic displays in decision making: the visual salience effect. *Journal of Behavioral Decision Making*, *3*, 247–262.
- Jarvenpaa, S. L., & Dickson, G. W. (1988). Graphics and managerial decision making: research based guidelines. *Communications of the ACM*, *31*, 764–774.
- Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist*, *58*, 697–720.
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*, 582–591.
- Kleinmuntz, D. N., & Schkade, D. A. (1993). Information displays and decision processes. *Psychological Science*, *4*, 221–227.
- Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology*, *76*, 972–987.
- Payne, J. B., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 534–552.
- Payne, J. W. (1982). Contingent decision behavior. *Psychological Bulletin*, *92*, 382–402.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: a constructive processing perspective. *Annual Review of Psychology*, *43*, 87–131.
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, England: Cambridge University Press.
- Payne, J. W., Bettman, J. R., & Luce, M. F. (1996). When time is money: decision behavior under opportunity-cost time pressure. *Organizational Behavior and Human Decision Processes*, *66*, 131–152.
- Rottenstreich, Y., & Tversky, A. (1997). Unpacking, repacking, and anchoring: advances in support theory. *Psychological Review*, *104*, 406–415.
- Russo, J. E., & Doshier, B. A. (1983). Strategies for multiattribute binary choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 676–696.
- Schkade, D. A., & Kleinmuntz, D. N. (1994). Information displays and choice processes: differential effects of organization, form, and sequence. *Organizational Behavior and Human Decision Processes*, *57*, 319–337.
- Shah, P., Freedman, E., & Vekiri, I. (In press). The comprehension of quantitative information in graphical displays. In P. Shah, & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking*. New York: Cambridge University Press.
- Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: signaling techniques for guiding the process of graph comprehension. *Journal of Educational Psychology*, *91*, 690–702.
- Simkin, D., & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, *82*, 454–465.
- Slooman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.
- Slovic, P., Griffin, D., & Tversky, A. (2002). Compatibility effects in judgment and choice. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 217–229). Cambridge, England: Cambridge University Press.
- Stone, E. R., Sieck, W. R., Bull, B. E., Yates, J. F., Parks, S. C., & Rush, C. J. (2003). Foreground: background salience: explaining the effects of graphical displays on risk avoidance. *Organizational Behavior and Human Decision Processes*, *90*, 19–36.
- Svenson, O. (1979). Process descriptions of decision making. *Organizational Behavior and Human Performance*, *23*, 86–112.

- 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. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, 79, 281–299.
- Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–98). Cambridge, England: Cambridge University Press.
- Tversky, A. & Koehler, D. J. (1994). Support theory: a nonextensional representation of subjective probability. *Psychological Review*, 101, 547–567.
- Windschitl, P. D., & Chambers, J. C. (2004). The dud-alternative effect in likelihood judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 198–215.
- Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: verbal versus numeric methods. *Journal of Experimental Psychology: Applied*, 4, 343–364.
- Windschitl, P. D., & Wells, G. L. (1998). The alternative-outcomes effect. *Journal of Personality and Social Psychology*, 73, 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 and Cognition*, 30, 469–477.
- Wright, P. L. (1974). The harassed decision maker: time pressures, distraction, and the use of evidence. *Journal of Applied Psychology*, 59, 555–561.

Authors' biographies:

Paul Windschitl is an associate professor of Psychology at the University of Iowa. His research interests include likelihood judgment, comparative judgment, social comparison, and egocentrism.

Zlatan Krizan is a graduate student at the University of Iowa. His research interests include wishful thinking, self-evaluation, social comparison, and likelihood judgment.

Authors' addresses:

Paul D. Windschitl and **Zlatan Krizan**, E11 SSH, Department of Psychology, University of Iowa, Iowa City, IA 52242, USA.