Journal of Behavioral Decision Making
J. Behav. Dec. Making, **16**: 107–126 (2003)
Published online 19 February 2003 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/bdm.437

The Influence of Reasons on Interpretations of Probability Forecasts

ANNETTE R. FLUGSTAD and PAUL D. WINDSCHITL*
University of Iowa, USA

ABSTRACT

When providing a probability estimate for an event, experts often supply reasons that they expect will clarify and support that estimate. We investigated the possible unintended influence that these reasons might have on a listener's intuitive interpretation of the event's likelihood. Experiments 1 and 2 demonstrated that people who read positive reasons for a doctor's probability estimate regarding a hypothetical surgery were more optimistic than those who read negative reasons for the identical estimate. Experiment 3 tested whether a doctor's failure forecast for a surgery would result in differing levels of pessimism when the potential risk was attributed to one complication that had a probability of 0.30 versus three complications that had a disjunctive probability of 0.30. Overall, the findings are consistent with the argument that a probability estimate, albeit numerically precise, can be flexibly interpreted at an intuitive level depending on the reasons that the forecaster provides as the basis for the estimate. Copyright © 2003 John Wiley & Sons, Ltd.

KEY WORDS probability; optimism; risk perception; reasons; forecasts; evaluability

When facing an important decision, people must often rely on an expert to provide the most informed estimate of the likelihood of a relevant outcome. For example, most people facing surgery have little idea how to determine their chances of a successful surgery, and therefore, need to rely on their doctor to provide the 'bottom-line' information about the likelihood that a surgery will be a success. Hence, it is critical that doctors and other experts provide likelihood information that is readily understood and not easily misinterpreted.

For experts and laypeople, the best form of bottom-line information would appear to be a numeric probability or relative-frequency estimate. A numeric estimate provides people with precise information on a standard metric that has a shared and externally-referenced interpretation. The weakness of the main alternative to numeric probability forecasts—verbal uncertainty expressions—is that their interpretations can be quite vague (see e.g. Beyth-Marom, 1982; Bryant & Norman, 1980; Budescu & Wallsten, 1985; Sutherland et al., 1991; see Budescu & Wallsten, 1995, for a review). An expression such as 'quite unlikely' might be interpreted as suggesting a 0.05 probability by one person and a 0.20 probability by another. Relatedly, the

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

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

meaning of verbal likelihood expressions can vary depending on context (Brun & Teigen, 1988; Budescu & Wallsten, 1995; Fox & Irwin, 1998; Mullet & Rivet, 1991; Wallsten et al., 1986; Weber & Hilton, 1990). The term 'likely' might be interpreted differently when it is used to describe the chances of a snowstorm in Minnesota than when it is used to describe the chances of a snowstorm in Missouri. For these and other reasons, many researchers have explicitly argued that physicians should, whenever possible, communicate uncertainty information, such as the likelihood that a surgery will be successful, with numeric values rather than verbal expressions (Marteau et al., 2000; Nakao & Axelrod, 1983; Welkenhuysen et al., 2001).

While noting the overwhelming support for the idea that verbal likelihood expressions are somewhat vague and susceptible to contrast effects, it is also important to note that numeric probability forecasts are not immune to interpretation problems. A forecaster's probability estimate, while being very precise in terms of a formal numerical system, is somewhat ambiguous in terms of how it should be intuitively interpreted (see Moxey & Sanford, 2000; Teigen & Brun, 1999, 2000; Windschitl et al., 2002; Windschitl & Weber, 1999). For example, learning that a speaker said that there is an 18% chance of X does not tell a listener whether the speaker is trying to convey a possibility, a doubt, or neither. In contrast, verbal expressions of likelihood tend to be less ambiguous in terms of direction; a speaker who uses the phrase 'small chance' is interpreted as expressing a possibility, albeit small, that something might happen, whereas a speaker who uses the phrase 'highly unlikely' is interpreted as expressing a doubt (see Moxey & Sanford, 2000; Teigen & Brun, 2000). More broadly, Windschitl et al. (2002) argued that any numeric probability—whether it is a communicated forecast, an internal belief regarding the objective likelihood of an event, or external information on which a belief is based—can be ambiguous from an intuitive perspective even though it is numerically precise.

This contention can be connected to work by Hsee and colleagues on the broader notion of evaluability (Hsee, 1996; Hsee et al., 1999). We suggest that an isolated numeric probability forecast is often difficult to evaluate and therefore does not have strong affective or intuitive implications (see also Kunreuther et al., 2001; Slovic et al., 2002). This makes probability forecasts susceptible to the influence of context. Learning that you have a 15% chance of experiencing a specific negative outcome may lead to alarm or complacency depending on contextual factors. Any context information that reduces the directional ambiguity of a forecaster's numeric estimate or enhances an audience's ability to meaningfully evaluate that numeric estimate relative to some standard will shape intuitive reactions.

Windschitl and Weber (1999) demonstrated how one type of context information can have an assimilation effect on people's intuitive responses to an expert's precise numeric forecast. In one of their studies, participants read about a woman who was told by her doctor that, because of a diagnosed blood condition, she would have a 30% chance of developing a disease related to malaria on an upcoming trip. Participants then indicated their certainty on a nonnumeric scale that she would develop the disease on her trip. Participants who read that the doctor's 30% estimate referred to a trip to India expressed greater certainty that she would get the disease than did participants who read that the 30% estimate referred to a trip to Hawaii. Results from studies using a variety of scenarios indicated that when respondents held an *a priori* association between the event in question and the described context (e.g. snow in the Colorado mountains), the context information would influence respondents' certainty about the event in the direction of assimilation (Windschitl & Weber, 1999).

Other recent experiments have demonstrated how context information of a different form can produce contrast rather than assimilation effects in the interpretations of numeric probability/frequency information (Windschitl et al., 2002). Participants in one experiment read about various diseases' prevalence rates for a target group (e.g. women) and a context group (e.g. men). The prevalence rates for a given disease within the target group were held constant (e.g. 12%), but the rates for the disease within the context group were manipulated (either 4% or 20%). Participants used the prevalence rate for the context group as an immediate comparison standard for interpreting the rate for the target group. Consequently, participants intuitively perceived the target group as more vulnerable to the disease when the prevalence rate for the context group was low (4%) rather than high (20%). In another study, Windschitl et al. demonstrated that these contrast-induced shifts in intuitive perceptions could create situations in which participants' beliefs about which of two events had a higher objective probability conflicted with their own intuitive expectations about which of the events was more likely.

To summarize, although interpretations of numeric probability expressions may be less vulnerable to context effects than are verbal probability expressions, they are not immune. The above-mentioned research demonstrated how context information that matches *a priori* associations in memory can yield assimilation effects (Windschitl & Weber, 1999) and context information that constitutes an immediate comparison standard can yield contrast effects (Windschitl et al., 2002).

In the present work, we investigated the potential influence of a different type of context information: the reasons that a forecaster cites for a probability estimate. Rarely do people receive a numeric probability forecast without some form of an accompanying explanation. To clarify and explain their forecasts, meteorologists show maps of pressure systems, stock market forecasters discuss various economic trends and signs, and physicians discuss test results, symptomology, and other factors with their patients. We suspected that in such cases, an audience would use these reasons or forms of supporting information as strong contextual cues for intuitively interpreting the expert's probability forecast. For example, we suspected that the optimism of a 37-year-old person will differ if he or she is told 'There is a 40% chance your joint will recover its full motion given you are under 40 years old' or 'There is a 40% chance your joint will recover its full motion given you are over 35 years old'. The former statement contains a positive cue given the situation under consideration, whereas the latter statement contains a negative cue. Similar to research showing that positively valenced primes lead to more favorable evaluations of ambiguous stimuli than do negative primes, we expected that a positive rather than a negative reason for a 40% estimate would lead to more favorable intuitive reactions to that intuitively ambiguous estimate (e.g. Bargh & Pietromonaco, 1982; Krosnick et al., 1992; Murphy & Zajonc, 1993). It is also the case that participants might use positive and negative reasons as a reflection of how the expert perceives the 40% estimate (e.g. unusually high relative to the norm), which would shape participants' reactions to the estimate. As we will discuss more in the general discussion, processes of these two types have also been proposed as explanations for attribute framing effects (see Levin et al., 1998; McKenzie & Nelson, 2002).

Our proposal is broadly related to, but distinct from, research on the influence of reasons for choices under uncertainty (see Shafir, 1993; Shafir et al., 1993; Tversky & Shafir, 1992). Shafir (1993), for example, investigated people's choices between two options—one richly described with both positive and negative attributes and another sparsely described. Whether people were told to select or reject an option, they tended to indicate the richly described option, which illustrates how the compatibility of a reason with the operation (selecting or rejecting) can influence choice.

It is important to emphasize that our focus is on the possibility that the reasons cited by an expert can influence people's intuitive optimism even when the forecaster indicated that the information underlying those reasons was already incorporated into, or formed the basis of, his or her numeric estimate. From the forecaster's perspective, the reasons he or she cites for a forecast should not be used by the audience to adjust the estimate. In other words, we focus on the impact of reasons cited by the forecaster for his or her *bottom-line* estimate. Although we expected that reasons offered for a bottom-line estimate would influence participants' optimism, we did not assume that participants would necessarily think that the experts' bottom-line estimate was wrong or in need of adjustment. Instead, we assumed that participants would generally believe that the experts' numeric estimate was the best estimate regarding objective numeric probability, but that their intuitive interpretation of the estimate would differ depending on the additional information that accompanied it.

OVERVIEW OF EXPERIMENTS

We conducted three experiments in which participants read a hypothetical scenario that suggested they were about to have surgery on a duct leading to their gall bladder. The scenario mentioned various pieces of information provided by their doctor, including a precise numeric estimate reflecting the doctor's 'bottom-line' assessment of the probability that the surgery would fail. The primary dependent measure asked participants to report what their intuitive optimism/pessimism would be regarding the surgery. In Experiment 1, we manipulated the primary valence of the reasons supporting the numeric forecast as well as the actual numeric forecast regarding failure. In Experiment 2, we held the failure forecast constant and manipulated participant's relative social-comparison status as well as the primary valence of the mentioned reasons. In Experiment 3 we again held the failure forecast constant, but we manipulated the reasons in a slightly different manner. Specifically, we manipulated the number, not the valence, of complications that could lead to an unsuccessful surgery. It is important to note that in these experiments, we did not directly manipulate the stated overall optimism/pessimism of the doctor (except in the forecast manipulation of Experiment 1). We simply inserted and manipulated information that a doctor might commonly mention in support of his or her assessment of a patient's risks related to surgery.

EXPERIMENT 1

As indicated in the above overview, participants in Experiment 1 read a hypothetical surgery scenario that included a doctor's bottom-line failure forecast as well as some additional information that essentially served as reasons for the failure forecast. The failure forecast (either 10% or 20%) and valence of the reasons (mostly 'good' risk-decreasing reasons or mostly 'bad' risk-increasing reasons) were manipulated in a between-subjects fashion. We predicted that despite the precision of the doctor's bottom-line probability forecast, its intuitive interpretation and the participants' resulting optimism would be heavily influenced by the general valence of the reasons provided by the doctor. We expected, for example, that a 20% failure probability that was supported by largely positive-valenced information would result in more optimism than a 20% failure probability that was supported by negative-valenced information. We manipulated the bottom line estimate (10% or 20%) so that we could compare the magnitude of a valence effect to the magnitude of an effect produced by a 10% difference in forecast probability.

Method

Participants and design

The participants were 126 students from Elementary Psychology classes at the University of Iowa, who participated to fulfill a research exposure component of the course. A 2×2 between-subjects design was used with reason valence (positive versus negative) crossed with failure forecast (10% versus 20%).

Procedure

The experimenter randomly assigned participants to read one of four versions of the main scenario. After participants read the scenario, they answered a set of dependent measures described below.

The scenario

The scenario instructed participants to imagine that they needed surgery to repair a duct leading to their gall bladder (see the Appendix for the full scenario). If the duct could not be repaired, removal of the gall bladder would become necessary. Removal of the whole gall bladder was described as a major surgery with its own complications and something that should be avoided. This was all to impress upon participants the importance of a successful surgery to repair the duct.

The scenario went on to describe five factors influencing the chances of having a successful surgery. Half of the participants read predominantly positive factors (4 positive and 1 negative) and half read predominantly negative factors (4 negative and 1 positive). An example of negative risk information included ‘your stomach lies low in your abdomen making it difficult to reach the duct needing repair’.[1] The doctor’s bottom-line failure forecast (either 10% or 20%) was then introduced in the following sentence of the scenario: ‘Given all these factors, the doctor tells you that the chance that your surgery will fail is X%.’ The wording of this bottom-line statement is important in that the sentence clearly indicates that the doctor’s estimate has presupposed the reasons that were described.

Dependent measures

The primary dependent measure was worded as follows: ‘Given this information, how pessimistic or optimistic would you feel about having a successful surgery?’ (1 = *extremely pessimistic*, 9 = *extremely optimistic*). Two other dependent measures assessed related constructs. The second question asked participants how much they would worry about the surgery failing (1 = *not worry at all*, 9 = *worry a lot*), and the final question asked participants how interested they would be in using an expensive and inconvenient medicine as a non-surgical alternative (1 = *not at all interested*; 9 = *extremely interested*).

Results and discussion

Table 1 presents the means and standard deviations for all three questions used in Experiment 1. The correlations among responses to these questions were statistically significant (all $ps < 0.05$) but small enough to consider the responses from these questions separately (the r’s absolute values ranged from 0.25 to 0.41).

The primary issue of interest was how participants’ optimism responses were influenced by the manipulations. As expected, an analysis of variance on optimism responses revealed a significant main effect for the

Table 1. Mean ratings in Experiment 1 by valence of reasons and failure forecast

	Failure rate					
	Low (10%)		High (20%)		Totals	
Valence of reasons	*M*	*SD*	*M*	*SD*	*M*	*SD*
Negative						
Optimism	5.9	1.6	5.5	1.5	5.7	1.5
Worry	6.2	2.1	6.1	1.9	6.1	2.0
Alternative	5.5	1.7	5.9	1.5	5.7	1.6
Positive						
Optimism	7.2	1.0	6.6	1.5	6.9	1.3
Worry	4.8	1.7	5.1	1.8	5.0	1.7
Alternative	4.9	1.8	4.8	2.0	4.9	1.9
Totals						
Optimism	6.6	1.5	6.1	1.6	6.1	1.5
Worry	5.5	2.1	5.6	1.9	5.5	2.0
Alternative	5.2	1.8	5.4	1.8	5.3	1.8

[1]We used reasons that would, to a typical participant, sound like plausible positive or negative factors regarding the described surgery. We were less concerned with whether these reasons were actually factors that would influence the likelihood of a real gall bladder surgery.

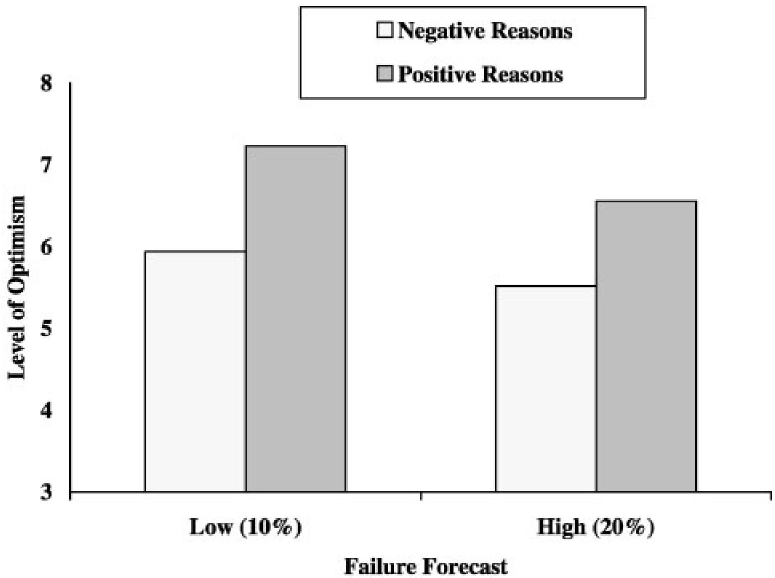

Figure 1. Mean optimism ratings from Experiment 1 as a function of the valence of reasons and failure forecast probability described in the scenario.

valence of reasons cited by the doctor, $F(1, 122) = 21.55$, $p < 0.001$. The failure-probability main effect was also significant, $F(1, 122) = 4.65$, $p < 0.05$, whereas the interaction term was not $F(2, 121) = 0.28$, $p > 0.05$. Figure 1 provides a graphical representation of these findings. A key conclusion evident from that representation and supported by the ANOVA is that participants who read largely positive reasons felt notably more optimistic than those who read largely negative reasons.

An ANOVA on responses to the worry question leads to a compatible conclusion. The main effect for valence of reasons was significant, $F(1, 122) = 12.43$, $p < 0.001$; participants who read largely positive reasons felt notably less worried than those who read largely negative reasons. The failure-probability main effect and the interaction were not significant, both $F\text{s} < 1$. Finally, an analysis of participants' reported interests in a non-surgical alternative also produced compatible results. The main effect for valence of reasons was significant, $F(1, 122) = 7.01$, $p < 0.01$; participants who read largely positive reasons for the doctor's forecast expressed less interest in an alternative than did participants who read largely negative reasons.

Overall, the results of Experiment 1 support the idea that, although an expert such as a surgeon may intend for a probability forecast to be interpreted as a 'bottom-line', people's feelings of optimism or vulnerability can be systematically influenced by the valence of reasons cited in support of the probability forecast. The magnitude of the valence effect was far from trivial. The effect size on the main dependent variable of optimism was 0.81, which is large according to conventional standards (Cohen, 1988). Perhaps more informative for evaluating the magnitude of the valence effect, we note that the influence of the valence manipulation was much stronger than that of the bottom-line failure probability ($d = 0.34$). In fact, it is interesting to note that the mean of the optimism responses from participants who read positive information about a 20% failure probability ($M = 6.6$) was actually 0.62 units higher than the mean from participants who read negative reasons for the objectively safer 10% failure probability ($M = 5.9$). This difference was not significant, $t(62) = 1.62$, $p = 0.11$, but the data pattern indicates that the effect of valence was not at all overshadowed by an effect of failure probability.

EXPERIMENT 2

Consistent with the work described in the introduction, we suspect even when participants in Experiment 1 believed that the surgeon's forecast was the best numeric estimate of failure, they nevertheless had distinct intuitive interpretations or reactions to the estimate—with positive (negative) reasons making the failure probability seem relatively small (large). While this explanation is relatively straightforward, a more thorough interpretation of Experiment 1 must involve a consideration of the possible role of presumed social-comparison status. When reading about a doctor's positive or negative reasons for a surgery's failure probability, it seems natural to assume that these reasons reflect something about whether the doctor believes the patient's health and specific characteristics are better or worse in comparison to others undergoing the same surgery. For example, if the doctor indicates to the patient 'you have very good blood flow', the patient is likely to assume that the doctor believes his or her blood flow is as good or better *than that of other patients*. A set of predominately positive reasons like this one could, consequently, suggest to the patient that his or her overall chances are better than others who undergo this surgery. Research by Klein (1997) suggests that this type of social comparison information—knowing your chances are better or worse than those of others—can influence reactions on behavioral, self-evaluative, and affective measures even when objective information predicts a different reaction. For example, in one of Klein's studies, participants who were told to imagine that they had a 60% chance of developing a health disorder and that this chance was lower than for the average person felt less disturbed than participants who imagined that they only had a 30% chance of developing the disorder but that this was a greater chance than for the average person. Hence, with respect to Experiment 1, it could be the case that the valence manipulation influenced perceived optimism through a simple social-comparison inference process. Specifically, participants used the reasons mentioned in the scenario as a cue to determining their social-comparison status, which then influenced their feelings of optimism.

In Experiment 2, we used a scenario similar to that of Experiment 1 and manipulated the valence of reasons, but we also manipulated the content of explicit information about social-comparison status. Although our discussion above highlights a possible mediational role for social comparison information, we did not design Experiment 2 to directly test this role with a standard mediational analysis. Rather, we tested whether valence of reasons would influence perceived optimism even when social-comparison or 'relative-risk' status was explicitly known—as would often be the case in real-world situations. Finding that valence of reasons influences perceived optimism even when 'relative-risk' status is known would suggest that participants did not simply use valence of reasons as a dichotomous relative-risk cue (better or worse than others), which in turn influenced optimism. More specifically, an additive effect of the two manipulations would suggest that valence of reasons can influence optimism independently of the presumed social-comparison status of a patient. The main alternative possibility of note was an interactive effect of the following type: the reasons manipulation would influence optimism when social-comparison status was not explicitly specified but would have little influence on optimism when social-comparison status was explicitly specified.

Another issue addressed by Experiment 2 was whether participants reading a hypothetical surgery scenario like those used here carefully consider the expert's numeric probability forecast. A rather uninteresting explanation for Experiment 1 would suggest that participants simply failed to read the numeric probability information contained in the scenario. Although we found this explanation to be highly improbable, it should be ruled out empirically. Hence, in Experiment 2, participants' memories for the numeric estimate were tested after they completed the main dependent measures.

Method

Participants and design

The participants were 240 students from Elementary Psychology classes at the University of Iowa, who participated to fulfill a research exposure component of the course. This was a 2×3 between-subjects design with valence of reasons (negative versus positive) crossed with social-comparison status (worse versus better versus none).

Procedure

Participants read one of six randomly assigned versions of a scenario describing a fictitious situation regarding a surgery. They then completed the set of dependent measures without access to this scenario information.

The scenario

The base scenario used in Experiment 2 was very similar to that used in Experiment 1. A given version of the scenario included either predominately positive or negative reasons (as in Experiment 1), and it included either downward comparison information, upward comparison information, or no comparison information. More specifically, downward-comparison versions included the following: 'Your doctor also tells you that compared to most people who need this type of surgery, your chances of a successful surgery are quite good.' Upward-comparison versions included the following: 'Your doctor also tells you that compared to most people who need this type of surgery, your chances of a successful surgery are not quite as good.' This sentence was omitted in the no-comparison condition. Finally all versions of the scenario ended with the following statement regarding the doctor's bottom-line estimate: 'He also gives you a numeric estimate about your chances. Specifically, he says there is a 15% chance of failure.'

Dependent measures

The measures were the same as in Experiment 1 with two exceptions. First, the worry question was followed by a new question: 'Of course needing gall bladder surgery is bad news. Aside from this, however, how would you feel overall about what your doctor told you?'($1 =$ *pretty bad*; $9 =$ *pretty good*). Second, a recall question, added at the end, asked participants to record the exact numeric estimate given by the doctor in the scenario.

Results

Table 2 presents the means and standard deviations for all four questions used in Experiment 2. The correlations among responses to these questions were statistically significant (all ps < 0.05) but again small enough to consider the responses from these questions separately. The absolute values of the correlations among the optimism, worry, and overall-feeling responses ranged from 0.41 to 0.61. The absolute values for correlations between these responses and the interest-in-nonsurgical-alternative responses were smaller, ranging from 0.19 to 0.25.

Figure 2 presents a graphical depiction of the pattern of means for the optimism responses, which were submitted to a 2×3 ANOVA. As in Experiment 1, the main effect for valence of reasons cited by the doctor was significant, $F(1, 234) = 35.74$, $p < 0.001$; the participants who read largely positive reasons felt better than those who read largely negative reasons. Also as expected, there was a significant main effect for social-comparison status, $F(2, 234) = 9.03$, $p < 0.001$. Follow-up comparisons revealed that participants in the worse-than-others condition felt more pessimistic about the surgery than participants in the no-comparison condition and the better-than-others condition (ps < 0.01); the difference in responses between the latter two groups was not significant ($p > 0.05$). Finally, the interaction between valence of reasons and social-comparison status was not significant, $F(2, 234) = 1.21$, $p = 0.30$. In other words, the reasons-valence manipulation had approximately the same impact on optimism in the conditions in which social-comparison status was explicitly mentioned versus the conditions in which comparison status was not explicitly mentioned.

Table 2. Mean ratings in Experiment 2 by valence of reasons and social comparison information

Valence of reasons	Social comparison status: Better than most		None		Worse than most		Totals	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Negative								
Optimism	5.6	1.4	5.7	1.6	4.6	1.6	5.3	1.6
Worry	6.3	2.0	6.6	1.6	7.1	1.4	6.7	1.7
Overall	5.4	1.4	4.7	1.7	4.3	1.6	4.8	1.6
Alternative	5.2	2.0	5.2	2.2	6.0	2.3	5.5	2.2
Positive								
Optimism	6.7	1.2	6.4	1.2	6.0	1.3	6.4	1.2
Worry	5.4	2.0	5.9	1.7	5.8	1.6	5.7	1.8
Overall	6.7	1.0	6.0	1.4	5.8	1.4	6.1	1.3
Alternative	4.3	2.4	5.3	2.6	5.6	2.2	5.1	2.4
Totals								
Optimism	6.1	1.4	6.1	1.4	5.3	1.6	5.8	1.5
Worry	5.9	2.1	6.2	1.7	6.5	1.6	6.2	1.8
Overall	6.0	1.4	5.3	1.7	5.1	1.7	5.5	1.6
Alternative	4.8	2.2	5.2	2.4	5.8	2.3	5.3	2.3

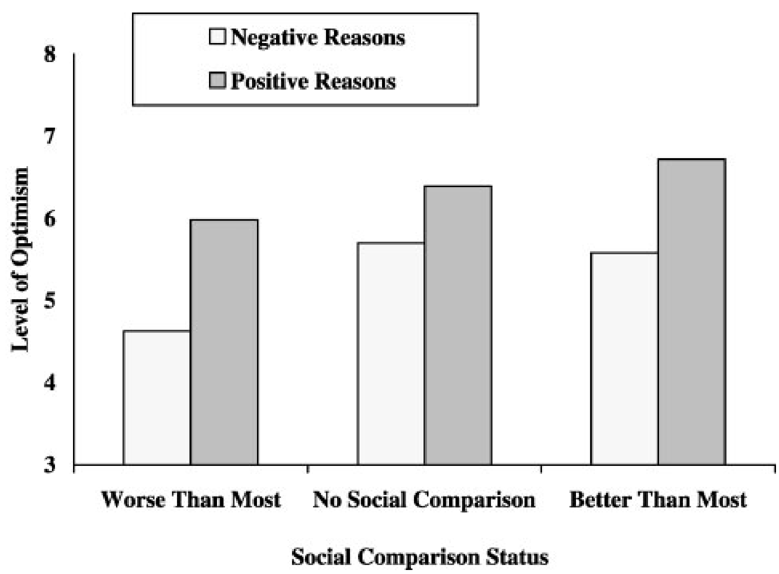

Figure 2. Mean optimism ratings from Experiment 2 as a function of the valence of reasons and social-comparison status described in the scenario.

Separate ANOVAs on responses to the worry and overall-feeling questions revealed patterns of results highly consistent with those described above. Specifically, the main effects for valence of reasons and for social comparisons status were significant ($p < 0.05$) for both questions, with the exception that the main effect for social-comparison status on worry only approached significance ($p = 0.08$). The interaction terms were nonsignificant for both analyses. The analysis of responses to the nonsurgical-alternative question produced somewhat different results. Unlike in Experiment 1, the reasons-valence main effect was not significant, $F(1, 234) = 1.64$, $p > 0.05$. However, there was a significant main effect for social-comparison status, $F(2, 234) = 4.46$, $p < 0.05$. Most important for the present purposes was the fact that the interaction between reasons and social-comparison status was not significant, $F(2, 234) = 0.90$, $p > 0.05$.

The overall set of findings from Experiment 2 suggests that the valence of reasons provided in support of a doctor's forecast can influence a person's optimism, worry, and overall feelings—even if he or she is explicitly informed of his or her social comparison status. This finding does not necessarily indicate that the impact of the valence of reasons cannot be partially mediated by inferred social-comparison status. However, the finding does not bode well for a full-mediation account that assumes that the overall valence of reasons mentioned by a doctor is utilized by people as a simple dichotomous cue as to their social-comparison status, which then influences their optimism and related reactions.

It is important to note that the influences of both the valence of reasons and social-comparison status were robust even though all participants read that the doctor's bottom-line estimate of failure probability was 15%. Based on previous findings mentioned in the introduction, we suggested that participants' intuitive reaction to the 15% estimate would be flexibly affected by the context in which that estimate was embedded. However, a different and rather uninteresting alternative interpretation would suggest that participants simply failed to attend to the 15% estimate. The data from the recall question, which asked participants to recall the doctor's failure forecast, helps to rule out this alternative explanation. Of the 240 participants, 236 (98.3%) correctly wrote '15%'. The data were analyzed again without the four incorrect cases, and the patterns of results remained essentially identical.

EXPERIMENT 3

Whereas Experiments 1 and 2 investigated how the valence of reasons for a failure forecast might influence intuitive optimism, Experiment 3 investigated the influence of a slightly different reasons-based factor. In Experiment 3, we manipulated the number of possible complications (three or one) that could require the surgery to be stopped. The doctor's bottom-line estimate (30%) was held constant. More specifically, some participants read a scenario version that indicated that because of a possible problem with bio-glue failing, a possible problem with a major blood vessel, and a possible problem with an enlarged duct, the overall failure probability was 30%. Other participants read scenario versions that mentioned only one of these possible complications, but those versions also indicated that the overall failure probability was 30% because of the possible complication. The key issue of interest in Experiment 3 was whether a description of three small risks that yield an overall risk of 30% would produce relatively more or less pessimism than would a description of one big 30% risk.

We expected that three small risks would produce more pessimism than would one big risk. People in the three-complications condition would receive three times the negative-valenced priming/cueing as would those in a one-complication condition, and we expected that the differential impact of this priming would not be neutralized by the equivalence in participants' knowledge about the forecaster's probability estimate (30% in both cases). This prediction has connections with two related lines of research. First, in their research on the numerosity heuristic, Pelham et al. (1994) demonstrated that when judging an amount or likelihood, people sometimes overuse the numerosity of relevant units as a cue (see related work by Fiedler & Armbruster, 1994). In one experiment, participants who briefly viewed a set of coins tended to estimate a greater overall monetary value when there were many rather than a few coins, even though the true monetary value was equivalent. In another experiment, participants who made a quick judgment about a person described as having 9 positive traits (1: artistic, 2: athletic, 3: clever, 4: coordinated, 5: creative, 6: energetic, 7: intelligent, 8: musical, 9: smart) provided more positive assessment that when those 9 traits were organized into three main traits (1: artistic, creative, musical; 2: athletic, coordinated, energetic; 3: clever, intelligent, smart). Pelham et al. (1994) suggest that a plausible underlying basis of the numerosity heuristic is that in natural environments, numerosity and quantity are likely to be highly correlated. Applied to the present experiment, this work would suggest that a 30% chance of failure might seem somewhat larger when the number is based on three separate 'units' of complications rather than just one; within medical settings, it seems likely that the number of possible complications is correlated—but only partially—with the likelihood that a surgery or treatment would fail.

Second, research on unpacking effects (see e.g. Fischhoff et al., 1978; Teigen, 1974; Tversky & Koehler, 1994) has demonstrated that a packed hypothesis (e.g. that the next winner of the Boston Marathon will be from Africa) will often be judged as having a lower probability than the sum of the judged probabilities of its unpacked components (e.g. that the next winner would be from Kenya, Ethiopia, South Africa, some other African nation). A key reason why unpacking increases the overall (summed) probability of a hypothesis is because it increases the activation and salience of various possibilities that might be forgotten or underappreciated in a packed representation. Although we did not technically use an unpacking manipulation in Experiment 3, we expected that, like unpacking, mentioning three possible complications rather than one would increase the overall activation of something akin to a negative-possible-outcomes category ('things that could go wrong'). Furthermore, assuming that the activation of concerns is closely linked to intuitive optimism, we suspected that intuitive optimism could shift as a function of number of complications even though the overall failure rate would be held constant.

Although we expected less optimism in the three-complications condition than in the one-complication condition, the reverse possibility was also quite plausible. When three complications are said to combine for a 30% overall risk, a person might tend to implicitly or explicitly decompose the 30% into three smaller percentages. If this is done, then perhaps no single risk seems very formidable. There is some intuitive rationale for viewing one large risk as more salient and worrisome than many small risks. There is also empirical evidence related to the alternative-outcomes effect that illustrates that when judging the likelihood of a focal outcome, several small nonfocal alternatives can sometimes have less impact on that judgment than does a single but strong nonfocal alternative (Windschitl & Wells, 1998). The alternative-outcome effect suggests, for example, that a raffle player holding three tickets in a raffle will tend to feel more confident about winning when there are three weak competitors (specifically, three other players who each hold two tickets) than when there is one strong competitor (specifically, one other player who holds six tickets).

Method

Participants, design, and procedure

Participants were 241 students from an Elementary Psychology course at the University of Iowa, who participated to fulfill a research exposure component of the course. They were assigned to read one of four possible versions of a surgery scenario that mentioned either an enlarged-duct complication, a major-vessels complication, a bio-glue complication, or all three of these complications. Because we were primarily interested in comparing responses of participants reading the three-complications version to those of participants reading any of the one-complication versions, we did not assign equal numbers of participants to each of the four versions. Instead, we constrained our random assignment so that more participants read the three-complications version than any one of the one-complication versions. Hence, 95 participants read the three-complications version, whereas 48, 49, and 49 participants read the enlarged-duct, major-vessels, and bio-glue versions, respectively. As in Experiment 2, the participants read one of the versions of the surgery scenario and then completed the set of dependent measures without access to the scenario.

Scenario

Each of the four versions of the scenario instructed participants to imagine that they needed surgery to repair a duct leading to their gall bladder. This was described as a major surgery and therefore their doctor did a thorough examination and the results highlighted one or more potential complications that could lead to the surgery being unsuccessful. In the major-vessels version, the doctor indicated that some major blood vessels were found close to the duct needing repair and may make completion of the surgery impossible. In the enlarged-duct version, the doctor indicated that one of the valves near the duct was enlarged and might require surgery to be stopped. In the bio-glue version, the doctor indicated that allergic testing revealed a potential allergy to the bio-glue required for the surgery. In the three-complications version, the doctor mentioned all of these potential complications. Finally all versions of the scenario contained a statement regarding the doctor's bottom-line estimate and what led to that estimate. For example, in the three-complications version (see the Appendix), it read 'Your doctor says that, based on an assessment of results from the physical exam, he could estimate the overall likelihood that the surgery would not be successfully completed. Specifically he says that there is a 30% overall chance that the surgery would not be successfully completed because of the possible interference from major blood vessels, the possibility of an enlarged duct, or the possibility of an allergic reaction to the bio glue. Your doctor then goes on to describe why these possibilities would require that surgery be stopped. He also explains how the results from the physical exam revealed each of these possibilities.'

Dependent measures
The dependent measures were the same as those in Experiment 2.

Results
Table 3 presents the means and standard deviations for all dependent measures. As was the case for Experiments 1 and 2, the correlations among responses to these dependent measures were statistically significant (all ps < 0.05) but small enough to consider the responses from these questions separately. The absolute values of the correlations among the optimism, worry, and overall-feeling responses ranged from 0.43 to 0.51. The absolute values for correlations between these responses and the interest-in-nonsurgical-alternative responses were smaller, ranging from 0.18 to 0.29.

Figure 3 presents a graphical depiction of the pattern of means for the optimism responses. We used a weighted contrast to test our hypothesis that the scenario version that described three complications would produce greater pessimism than would the versions that mentioned only one complication. Specifically, the three-complications condition, enlarged-duct condition, major-vessels condition, and bio-glue condition were assigned weights of $+3$, -1, -1, and -1, respectively. This contrast test was significant, $t(236) = 3.09$, $p < 0.01$, thus providing support for our hypothesis. It is also important to note the results of pairwise comparisons between the conditions. None of the pairwise comparisons between responses from the enlarged-duct condition, major-vessels condition, or bio-glue condition were significant (all ps > 0.40).

Table 3. Mean ratings in Experiment 3 by the scenario version

	Complications in scenario version							
	Enlarged duct[a]		Major vessel[b]		Bio glue[c]		All three[d]	
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Optimism	5.0	1.7	4.8	1.9	4.9	1.7	4.2	1.7
Worry	6.7	1.7	6.6	2.0	7.1	1.7	7.0	1.8
Overall	4.4	1.5	4.2	1.9	3.9	1.6	3.6	1.5
Alternative	6.6	2.2	6.8	2.4	6.7	2.3	6.6	2.2

[a]$n = 49$; [b]$n = 49$; [c]$n = 47$; [d]$n = 95$.

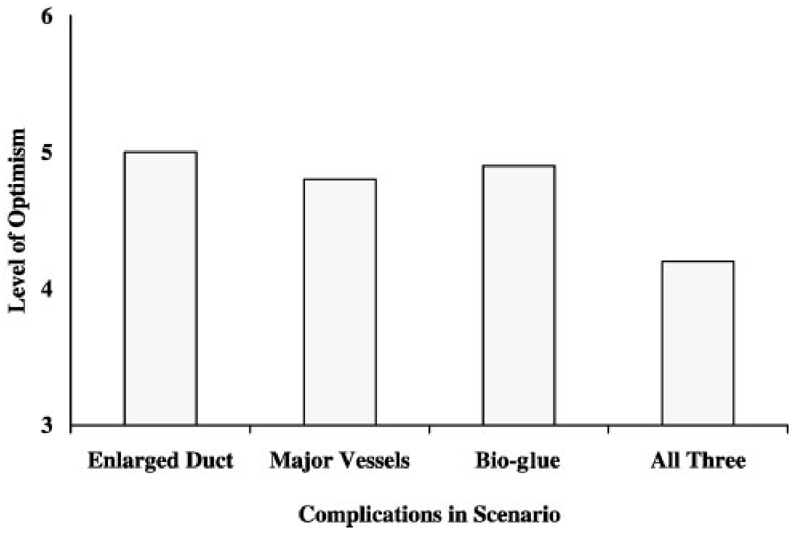

Figure 3. Mean optimism ratings from Experiment 3 as a function of the complications described in the scenario.

However, as predicted, the responses from the three-complications condition were significantly different from those in the enlarged-duct and bio-glue conditions (both $p\text{s} < 0.05$); the difference between the three-complications and major-vessels conditions was borderline ($p = 0.055$).[2]

Unlike for the optimism question, participants did not report significantly different levels of worry in the three-complications condition as compared to the one-complication conditions. More specifically, none of the weighted contrasts or pairwise comparisons conducted on the worry data were significant (all $p\text{s} > 0.05$). The results for the overall-feelings question, however, look much the same as those for the optimism question. As predicted, the weighted contrast (+3, −1, −1, −1) was significant, $t(236) = 2.69$, $p < 0.01$. None of the pairwise comparisons between the enlarged-duct condition, major-vessels condition, or bio-glue condition were significant (all $p\text{s} > 0.15$). The responses from the three-complications condition were significantly different from those in the enlarged-duct and major-vessels conditions (both $p\text{s} < 0.05$); the comparison with the bio-glue condition was directional but not significant ($p = 0.25$). Finally, as was the case for the worry data, none of the weighted contrasts or pairwise comparisons conducted on the responses to the nonsurgical-alternative question were significant (all $p\text{s} > 0.05$).

As in Experiment 2, participants were asked to recall the numeric estimate given by the doctor to determine if the effects on the intuitive vulnerability measures could be due to an inaccurate recall of the numeric probability. In Experiment 3, 90% of the participants (216 out of 240) correctly recalled the doctor's bottom-line estimate. All of the above analyses were re-run excluding data from participants who did not correctly recall the bottom-line estimate. The pattern of results remained essentially the same as what we've already described.

In sum, although the data from the worry and nonsurgical-alternative question did not significantly differ between the three-complications and one-complication conditions, the data from the optimism and overall feeling questions provide clear support for our hypothesis. The number of possible complications supporting

[2]This pattern of results rules out the possibility that the weighted contrast was significant only because participants were reacting in an especially negative way to only one of the complications that were described in the three-complications version. This would deflate the optimism of the three-complications group, but only deflate the optimism of one of the one-complication groups. This possibility could have resulted in a spurious effect for the 3, −1, −1, −1 contrast (see Abelson, 1996; Petty et al., 1996).

a failure probability has an impact on how intuitively optimistic people feel about the possibility of failure, even when the failure probability is known and held constant. Recall that a plausible alternative hypothesis was that the impact of one big risk would be greater than the combined impact of three smaller risks. However, the direction of the results indicate that the impact of one big risk was in fact smaller than the impact of three smaller risks, a finding that is consistent with research on the numerosity heuristic (Pelham et al., 1994). Our results are also broadly consistent with research on unpacking effects. Similar to other accounts for unpacking effects, we assume that listing three complications rather than one increases the overall salience and accessibility of risks. Experiment 3 leads us to speculate that classic unpacking manipulations might not only influence respondents' judgments or estimation of probability (see e.g. Fischhoff et al., 1978; Teigen, 1974; Tversky & Koehler, 1994), but also their interpretation of stated probability forecasts or known probabilities.

GENERAL DISCUSSION

This research illustrates how a forecaster's bottom-line probability estimate regarding an outcome might not serve as the bottom-line for the audience. In all three experiments, the reasons that a doctor cited for a bottom-line estimate influenced the intuitive optimism of the respondents separately from the value of the bottom-line estimate itself. In Experiments 1 and 2, the key manipulation was the valence of the reasons cited by the doctor (either predominantly negative or positive), whereas in Experiment 3 the key manipulation involved the quantity of negative reasons cited by the doctor.

What causal explanations can or cannot account for the effects detected in these experiments? One account that was ruled out in Experiments 2 and 3 was the possibility that participants simply failed to sufficiently process the bottom-line estimate provided by the doctor. In those experiments, 98.3% and 90% of the participants, respectively, were able to correctly recall the bottom-line estimate from memory.

A second account that was not consistent with results from Experiment 2 suggested that the valence of reasons cited by the doctor served as a simple dichotomous cue as to whether participants should assume that their chances during surgery would be worse or better than most other patients. In Experiment 2, the effect of the valence-of-reasons manipulation was just as strong when social-comparison status was known than when it was unknown. Also, in Experiment 3, the valence of the reasons cited was always negative (one versus three complications), which makes it difficult, albeit not impossible, to explain how reasons could serve as a cue to social-comparison status.

A third causal explanation that is possible but seems unlikely to account for the full results suggests that participants assumed that the reasons cited by the doctor constituted new information that the doctor had not already considered and presupposed when making his or her numeric probability estimate. Perhaps, for example, the participants interpreted the bottom-line estimate as a statement of the typical failure rate. In other words, perhaps some participants thought something like: 'I would normally have a 20% chance of failure, but the doctor says I have good blood flow, so the failure probability must be lower than 20%.' While we cannot rule out this type of thinking for each and every participant, we suspect that very few, if any, participants used this interpretation. Our statements of the bottom-line estimates were carefully constructed to make clear that the doctor's forecast presupposed all of the information discussed in the scenarios.

A fourth possible explanation suggests that participants found fault with the doctor's numeric probability estimate. Perhaps participants assumed that the doctor's bottom-line estimate was not the best estimate of objective probability, and instead generated a failure probability value based on their knowledge of surgery and the valenced reasons mentioned in the scenario. This explanation, however, seems implausible given that gall bladder surgery is an event for which our participants would have little or no expertise. Hence, it seems unlikely that participants would reject an expert's assessment and replace it with their own naïve assessment. Although we suggest that the main results of our experiments are not likely due to misperceptions of the doctor's bottom-line forecast or to replacements of the doctor's bottom-line statements, we also believe that future research should more directly test the possible influence of such misperceptions and replacements.

The explanation that we described in the introduction and that we believe is most plausible assumes that participants accepted the doctor's forecast as the best numeric estimate of the probability of failure. However, a forecaster's probability estimate, despite being very precise in terms of a formal numerical system, is ambiguous in that it does not carry with it direct implications for a person's more intuitive reactions such as gut-level optimism (see Windschitl et al., 2002; see also Moxey & Sanford, 2000; Teigen & Brun, 1999, 2000). Hence, the reasons offered by the forecaster in the present studies served as cues or primes that influenced the interpretation of the intuitively ambiguous probability estimate. Our findings add to a growing list of findings demonstrating how contextual information can affect the interpretation of a precise numeric probability value (Windschitl et al., 2002; Windschitl & Weber, 1999). More generally, these findings are consistent with the notion that people may hold a belief about the best numeric estimate of objective probability for an event but simultaneously hold a distinct intuitive expectation regarding the event (see e.g. Windschitl, 2000; Windschitl & Wells, 1996, 1998), and this notion can be connected to broader dual-processing proposals (e.g. Epstein, 1994; Sloman, 1996; see also discussion by Windschitl et al., 2002).

The priming/cueing processes that we suggested might underlie the effects of forecasters' reasons are related to processes proposed to account for attribute-framing effects. In attribute framing, an object or event is evaluated more favorably when a key attribute is framed in a positive (e.g. % correct, % wins) rather than negative (e.g. % incorrect, % loses) way (see the review by Levin et al., 1998). Levin and his colleagues have proposed that positive or negative labels serve as primes that tend to evoke valence-consistent associations regarding the target object, which ultimately affect the evaluation of the object (see e.g. Levin & Gaeth, 1988; Levin et al., 1998, 2002). McKenzie and Nelson (2002) recently proposed that a frame implicitly communicates the perspective of the speaker, which in turn could be used by a listener as an interpretational cue. For example, a listener who hears a speaker say a glass is half full is more likely to infer that the glass was recently empty rather than full. As we stated in the introduction, we suspect that reasons cited by a forecaster could serve as both a prime that triggers valenced associations and as a cue that listeners use to make inferences (e.g. about a forecaster's implicit perspective); both of these processes could ultimately play a role in shaping their intuitive optimism.[3]

THE ROLE OF THE FORECAST FORMAT ON INTERPRETATIONS

One issue not addressed in the present studies was whether the format of a forecast might play a moderating role on the types of context effects we investigated. For example, would context have the same influence when a doctor tells a patient that his or her surgery has a 10% chance of failure than when a doctor tells a patient that, for cases like his or hers, the surgery will fail 10 out of 100 times? Some researchers have argued that people are generally more capable when processing relative frequency information rather than probability information (e.g. Gigerenzer, 1994), and others have noted differences in how people respond to equivalent information represented as frequencies versus probabilities (e.g. Slovic et al., 2000). Only additional research can determine whether format would play a moderating role in the present paradigm. It is worth noting, however, that in a project on context effects that was described earlier (Windschitl et al., 2002), context effects were detected not only in cases when disease prevalence rates were described as

[3]If we examine the cueing proposal with respect to Experiment 2, that proposal would seem to predict that the reasons cited by the doctor would implicitly inform participants about whether their chances were better or worse than average. The fact that the valence-of-reasons effect was robust even when social comparison status was explicitly stated does not bode well for that type of cueing explanation of those particular results. However, we do not view this finding as a sign that people rarely or never use a forecaster's reasons to make inferences about his or her implicit views or to make inferences about whether a probability forecast is higher or lower than typical.

percentages (e.g. 12% of women and 4% of men), but also in cases when prevalence rates were described as relative frequencies (1 in 900 European Americans and 1 in 200 African Americans). We suspect, therefore, that communicating the risk of a surgery failure in terms of relative frequencies rather than a single-event probability would not preclude context effects of the sort studied in this work.

IMPLICATIONS

Like framing effects detected in medical-decision contexts (see e.g. Marteau, 1989; McNeil et al., 1982; Wilson et al., 1987), the effects detected in the present work have potentially serious implications for how doctors and other experts communicate uncertainty. Experts may often unwittingly provide one-sided support for a bottom-line probability estimate. For example, if a doctor has a patient whose failure probability for a surgery is slightly higher than average, the doctor may provide his or her bottom-line estimate and primarily describe reasons why this estimate is high and fail to mention any reasons why the estimate is not even higher. Another reason why an expert might provide one-sided support for a bottom-line estimate is if he or she suspects that the bottom-line estimate will be higher or lower than what is expected by the audience. In these situations, an audience may get what could be called a double dose of the same news—the news related to the numeric value itself but also the one-sided reasons for the value. Sometimes this double dose is delivered by an expert strategically, in hopes of having a strong impact on the audience, but sometimes this double dose might have an effect that was unanticipated.

Relatedly, Experiment 3 clearly shows that mentioning more potential reasons for failure (e.g. complications) can increase pessimism about the success/occurrence of the target event, despite equivalence in the overall failure probability. Extending this finding only slightly, we suspect that as a forecaster unpacks a category of reasons for failure (e.g. complications involving a restriction of blood circulation) into less likely but more numerous subcomponents of this category ('X, Y, and Z are all possible circulation problems'), an audience will become more pessimistic about the success/occurrence of the target event. While a forecaster might assume that providing a bottom-line estimate and/or emphasizing the improbable nature of any one impediment should minimize an audience's overall concern, Experiment 3 provides reasons to question this assumption. Additional research would be required to determine how a forecaster might unpack a category of possible impediments—hence being more descriptive for an audience—without unduly inflating pessimism at an intuitive level.

One limitation of this work is that we did not test for context effects in situations involving patients who were in fact considering surgery. Future research should determine whether the magnitude of these context effects is enhanced, similar, or reduced in such conditions. In these three laboratory studies involving hypothetical scenarios, the influence of context was quite robust. The effects sizes (Cohen's d) for the influence of the reasons manipulations on the optimism question in Experiments 1, 2, and 3, were 0.81, 0.75, and 0.41, respectively. Perhaps more informative is that fact that in Experiment 1, the valence-of-reasons manipulation had a stronger effect than did the manipulation of the bottom-line failure estimate from 10 to 20%. On one hand, one might expect that in a real-world situation, a patient would be highly motivated to process all information extremely carefully and deliberately. Hence, the observed context effects might be diminished because relative to our participants, real patients might do a better job of separating the bottom-line estimate from the context information, or they might keep their intuitive perceptions in close correspondence with their belief about objective probability. On the other hand, a more careful and deliberative processing of the information might actually increase the influence of the reasons cited by the doctor. A real patient who thinks at length or dwells on ways in which a surgery could go wrong might actually lose some interest in the rather abstract numeric probability of failure. It is also noteworthy that the contextual information that we manipulated in these experiments may be relatively mild compared to the contextual information and cues that might be present in nonhypothetical cases. We did not directly manipulate the stated optimism/pessimism of the doctor, nor did we manipulate or even mention recommendations made by the doctor. Facial expressions and cues from body language that may be influential in a nonhypothetical case were not explored in the present study, but are potential factors to be considered in future research with actual patients.

Finally, we note that although our methods were only a general approximation of how patients might receive likelihood information regarding surgery, people often encounter other types of likelihood-relevant information in a fashion quite similar to how the information was encountered by our research participants. Specifically, newspaper stories and other media stories about uncertain events (e.g. product failures, acquiring an illness, vaccine reactions) often include base-rate estimates or forecaster's bottom-line estimates that are supported by reasons. Hence, there are many applied situations in which the reasons that experts cite in support for their bottom-line estimate have the potential to have unintended consequences. For the experts providing these bottom lines estimates, we have two key messages: (1) Be aware that your bottom-line estimates *per se* can be quite ambiguous in the mind of an audience, and that the audience will use information accompanying that estimate in order to make intuitive sense of it; (2) Give careful consideration as to whether you should provide one-sided support for a bottom-line estimate. At times there may be good reasons to provide only information for why an estimate is high or why it is low, but there may also be times for which one-sided arguments can have robust and unwanted influences on uncertainty.

APPENDIX

Below are selected versions of scenarios used in Experiments 1 and 3. For these scenarios, we used reasons/complications that would, to a typical participant, sound like plausible positive or negative factors regarding the described surgery.

The positive-valenced scenario used in Experiment 1

Imagine that you need to have surgery to repair the duct leading to your gall bladder. If it is unsuccessful, your gall bladder will need to be completely removed. Removing your gall bladder is a major surgery and can lead to other complications, which your doctor wants to avoid. The doctor tells you that there are some major factors influencing the chances of a successful repair of the duct. One piece of good news is that you have large veins, which may provide good blood flow to the duct. This will make the surgery easier. You also have normal blood pressure, which decreases the complexity of the surgical procedure. However, your stomach lies low in your abdomen making it difficult to reach the duct needing repair. On the other hand, there is no blockage of the duct, shortening of the length of the surgery. The less amount of time the surgery takes, the more likely it is that the repair of the duct will be successful. Finally, the doctor tells you that your liver produces a small amount of bile, which should not be a problem in the surgery. Given all these factors, the doctor tells you that the chance that your surgery will fail is *X*%.

The negative-valenced scenario used in Experiment 1

Imagine that you need to have surgery to repair the duct leading to your gall bladder. If it is unsuccessful, your gall bladder will need to be completely removed. Removing your gall bladder is a major surgery and can lead to other complications, which your doctor wants to avoid. The doctor tells you that there are some major factors influencing the chances of a successful repair of the duct. One piece of bad news is that you have small veins, which may obstruct blood flow to the duct. This will make the surgery more difficult. You also have high blood pressure, which increases the complexity of the surgical procedure. However, your stomach lies high in your abdomen making it easier to reach the duct needing repair. On the other hand, there is blockage of the duct, which could double the length of the surgery. The longer the surgery takes, the less likely it is that the repair of the duct will be successful. Finally, the doctor tells you that your liver produces a large amount of bile, which may interfere with the surgery. Given all these factors, the doctor tells you that the chance that your surgery will fail is X%.

The three-complications scenario used in Experiment 3

Imagine that you have a medical problem with a duct that leads to your gall bladder. The main option for dealing with this problem is a major surgery. Before deciding on whether to have the surgery, you first undergo a thorough physical evaluation. After the evaluation, your doctor meets with you to discuss what the results were and how the results affect your chances of having a successful surgery. The doctor tells you that the pre-surgery evaluation suggested caution about three potential complications during surgery. First, from preliminary scans it appears that some major blood vessels lie close to the areas involved in the operation. There is a possibility that the location of these vessels will make completion of the surgery impossible. Second, preliminary scans indicate that there is a possibility that one of the valves near the duct will be enlarged, which would require that the surgery be stopped. Third, allergy testing revealed that you may have an allergic reaction to the bio-glue that is required for this surgery. If the surgeon sees signs of a reaction, the surgery will be halted. Your doctor says that, based on an assessment of results from the physical exam, he could estimate the overall likelihood that the surgery would not be successfully completed. Specifically, he says that there is a 30% overall chance that the surgery would not be successfully completed because of the possible interference from major blood vessels, the possibility of an enlarged duct, or of an allergic reaction to the bio glue. Your doctor then goes on to describe why these possibilities would require that the surgery be stopped. He also explains how the results from the physical exam revealed each of these possibilities.

ACKNOWLEDGMENTS

We thank Irwin Levin for his helpful comments on an earlier draft of this paper. This work has benefited from support by Grant SES 99-11245 from the National Science Foundation.

REFERENCES

Abelson, R. P. (1996). Vulnerability of contrast tests to simpler interpretations: an addendum to Rosnow and Rosenthal. *Psychological Science, 7*, 242–246.

Bargh, J. A., & Pietromonaco, P. (1982). Automatic information processing and social perception: the influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology, 43*, 437–449.

Beyth-Marom, R. (1982). How probable is probable? A numeric translation of verbal probability expressions. *Journal of Forecasting, 1*, 257–269.

Brun, W., & Teigen, K. H. (1988). Verbal probabilities: ambiguous, context-dependent, or both? *Organizational Behavior and Human Decision Processes, 41*, 390–404.

Bryant, G. D., & Norman, G. R. (1980). Expressions of probability: words and numbers. *The New England Journal of Medicine, 302*, 411.

Budescu, D. V., & Wallsten, T. S. (1985). Consistency in interpretation of probabilistic phrases. *Organizational Behavior & Human Decision Processes, 36*, 391–405.

Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: general principles and empirical evidence. *The Psychology of Learning and Motivation, 32*, 275–318.

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

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*, 709–724.

Fiedler, K., & Armbruster, T. (1994). Two halfs may be more than one whole: category-split effects on frequency illusions. *Journal of Personality and Social Psychology, 66*, 633–645.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception & Performance, 4*, 330–344.
Fox, C. R., & Irwin, J. R. (1998). The role of context in the communication of uncertain beliefs. *Basic and Applied Social Psychology, 20*, 57–70.
Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright, & P. Ayton (Eds.), *Subjective probability* (pp. 130–161). Chichester: John Wiley & Sons Ltd.
Hsee, C. (1996). The evaluability hypothesis: an explanation of preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 46*, 247–257.
Hsee, C. K., Loewenstein, G. F., Blount-Lyons, S., & Bazerman, M. H. (1999). Preference reversals between joint and separate evaluations of options: a review and theoretical analysis. *Psychological Review, 125*, 576–590.
Klein, W. M. (1997). Objective standards are not enough: affective, self-evaluative, and behavioral responses to social comparison information. *Journal of Personality and Social Psychology, 72*, 763–774.
Krosnick, J. A., Betz, A. L., Jussim, L. J., & Lynn, A. R. (1992). Subliminal conditioning of attitudes. *Personality and Social Psychology Bulletin, 18*, 152–162.
Kunreuther, H., Novemsky, N., & Kahneman, D. (2001). Making low probabilities useful. *The Journal of Risk and Uncertainty, 23*, 103–120.
Levin, I. P., & Gaeth, G. J. (1988). Framing of attribute information before and after consuming the product. *Journal of Consumer Research, 15*, 374–378.
Levin, I. P., Gaeth, G. J., Schreiber, J., & Lauriola, M. (2002). A new look at framing effects: distributions of effect sizes, individual differences, and independence of types of effects. *Organizational Behavior and Human Decision Processes, 88*, 411–429.
Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: a typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*, 149–188.
MacNeil, B. J., Pauker, S. G., Sox, H. C., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *New England Journal of Medicine, 306*, 1259–1262.
Marteau, T. M. (1989). Framing of information: its influence upon decisions of doctors and patients. *British Journal of Social Psychology, 28*, 89–94.
Marteau, T. M., Saidi, G., Goodburn, S., Lawton, J., Michie, S., & Bobrow, M. (2000). Numbers or words? A randomized controlled trial of presenting screen negative results to pregnant women. *Prenatal Diagnosis, 20*, 714–718.
McKenzie, C. R. M., & Nelson, J. D. (2002). What a speaker's choice of frame reveals: reference pointes, frame selection, and framing effects. *Psychonomic Bulletin and Review* (conditionally accepted).
Moxey, L. M., & Sanford, A. J. (2000). Communicating quantities: a review of psycholinguistic evidence of how expressions determine perspectives. *Applied Cognitive Psychology, 14*, 237–255.
Mullet, E., & Rivet, I. (1991). Comprehension of verbal probability expressions in children and adolescents. *Language & Communication, 11*, 217–225.
Murphy, S. T., & Zajonc, R. B. (1993). Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures. *Journal of Personality and Social Psychology, 64*, 723–739.
Nakao, M. A., & Axelrod, S. (1983). Numbers are better than words. *The American Journal of Medicine, 74*, 1061–1065.
Pelham, B. W., Sumarta, T. T., & Myaskovsky, L. (1994). The easy path from many to much: the numerosity heuristic. *Cognitive Psychology, 26*, 103–133.
Petty, R. E., Fabrigar, L. R., Wegener, D. T., & Priester, J. R. (1996). Understanding data when interactions are present or hypothesized. *Psychological Science, 7*, 247–252.
Shafir, E. (1993). Choosing versus rejecting: why some options are both better and worse than others. *Memory & Cognition, 21*, 546–556.
Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition, 49*, 11–36.
Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*, 3–22.
Slovic, P., Finucane, M., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich, D. Kahneman, & D. Griffin (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). Cambridge: Cambridge University Press.
Slovic, P., Monahan, J., & MacGregor, D. G. (2000). Violence risk assessment and risk communication: the effects of using actual cases, providing instruction, and employing probability versus frequency formats. *Law and Human Behavior, 24*, 271–295.
Sutherland, H. J., Lockwood, G. A., Tritchler, D. E., Sem, F., Brooks, L., & Till, J. E. (1991). Communicating probabilistic information to cancer patients: is there 'noise' on the line? *Behavioral Science and Medicine, 32*, 725–731.
Teigen, K. H. (1974). Overestimation of subjective probabilities. *Scandinavian Journal of Psychology, 15*, 56–62.

Teigen, K. H., & Brun, W. (1999). The directionality of verbal probability expressions: effects on decisions, predictions, and probabilistic reasoning. *Organizational Behavior and Human Decision Processes, 80*, 155–190.
Teigen, K. H., & Brun, W. (2000). Ambiguous probabilities: when does $p = 0.3$ reflect a possibility, and when does it express a doubt? *Journal of Behavioral Decision Making, 13*, 345–362.
Tversky, A., & Koehler, D. J. (1994). Support theory: a nonextensional representation of subjective probability. *Psychological Review, 101*, 547–567.
Tversky, A., & Shafir, E. (1992). The disjunction effect in choice under uncertainty. *Psychological Science, 3*, 305–309.
Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory & Language, 25*, 571–587.
Weber, E. U., & Hilton, D. J. (1990). Contextual effects in the interpretations of probability words: perceived base rate and severity of events. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 781–789.
Welkenhuysen, M., Evers-Kiebooms, G., & d'Ydewalle, G. (2001). The language of uncertainty in genetic risk communication: framing and verbal versus numerical information. *Patient Education and Counseling, 43*, 179–187.
Wilson, D. K., Kaplan, R. M., & Schneiderman, L. J. (1987). Framing of decision selections of alternatives in health care. *Social Behavior, 2*, 51–59.
Windschitl, P. D. (2000). The binary additivity of subjective probability does not indicate the binary complementarity of perceived certainty. *Organizational Behavior & Human Decision Processes, 81*, 195–225.
Windschitl, P. D., 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., & Weber, E. U. (1999). The interpretation of 'likely' depends on the context, but '70%' is 70%—right? The influence of associative processes on perceived certainty. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1514–1533.
Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: verbal versus numeric methods. *Journal of Experimental Psychology: Applied, 2*, 343–364.
Windschitl, P. D., & Wells, G. L. (1998). The alternative-outcomes effect. *Journal of Personality and Social Psychology, 75*, 1411–1423.

Authors' biographies:
Annette Flugstad is a graduate student at the University of Iowa. Her research interests include heuristic processing, subjective probability judgments, and prejudice reduction.

Paul Windschitl is an Assistant Professor of Psychology at the University of Iowa. His research interests include perceptions of certainty, subjective probability judgments, and comparative social judgments.

Authors' address:
Annette Flugstad and **Paul Windschitl**, Department of Psychology, University of Iowa, Iowa City, IA 52242, USA.

