Risk Analysis, Vol. 29, No. 10, 2009 DOI: 10.1111/j.1539-6924.2009.01279.x

The Use of Narrative Evidence and Explicit Likelihood by Decisionmakers Varying in Numeracy

Nathan F. Dieckmann,* Paul Slovic, and Ellen M. Peters

Decisionmakers are often presented with explicit likelihood assessments (e.g., there is a 10% chance that an attack will occur over the next three months) and supporting narrative evidence in forecasting and risk communication domains. Decisionmakers are thought to rely on both numerical and narrative information to the extent that they perceive the information to be diagnostic, accurate, and trustworthy. In two studies, we explored how lay decisionmakers varying in numeracy evaluated and used likelihood assessments and narrative evidence in forecasts. Overall, the less numerate reported higher risk and likelihood perceptions. In simple probabilistic forecasts without narrative evidence, decisionmakers at all levels of numeracy were able to use the stated likelihood information, although risk perceptions of the less numerate were more affected by likelihood format. When a forecast includes narrative evidence, decisionmakers were better able to use stated likelihood in a percentage as compared to frequency or verbal formats. The more numerate used stated likelihood more in their evaluations whereas the less numerate focused more on the narrative evidence. These results have important implications for risk analysts and forecasters who need to report the results of their analyses to decisionmakers. Decisionmakers varying in numerical ability may evaluate forecasts in different ways depending on the types of information they find easiest to evaluate.

KEY WORDS: Forecasting; intelligence analysis; numeracy; risk communication; risk perception

There are many risk communication and forecasting contexts in which decisionmakers are presented with narrative evidence or other contextual information along with assessments of likelihood or risk. Different kinds of contextual information can accompany a likelihood assessment. For instance, anecdotal information about another person's experience with a medical treatment or event can be included when providing a likelihood assessment. Forecasters can also present reasons, justification, or logic supporting a likelihood assessment. In the legal domain, for instance, an expert witness may present supporting evidence in the form of a narrative and conclude with an opinion that there is an x% chance that the target event occurred. An intelligence analyst may prepare a report that includes both a discussion outlining his or her interpretation of the weight and inferential force of the evidence concerning an event along with numeric assessments of likelihood. This latter type of contextual information is called narrative evidence in this article. In these forecasting situations, decisionmakers will need to assess how much to trust the analysis of the expert and how to judge the overall risk of the forecasted event.

Research on information presentation and use has sometimes been framed as pitting "narrative" versus "numerical" information. In reality, decisionmakers likely rely on both numerical and narrative information to the extent that they perceive the

Decision Research, & University of Oregon Eugene, OR, USA.
* Address correspondence to Nathan F. Dieckmann, Decision Research, 1201 Oak Street, Eugene, OR 97401, USA; tel: 1-541-485-2400; fax: 1-541-485-2403; ndieckmann@decisionresearch.org.

information to be diagnostic and accurate. However, relatively few studies have explored judgments when both explicit likelihood and other contextual information is available to the judge. One consistent finding from the work that has been done is that contextual information accompanying a likelihood assessment can have a large effect on the interpretation of that assessment.[1–5] These findings can be connected to the evaluability work of Hsee and colleagues,[6] in that "an isolated numerical probability forecast is often difficult to evaluate and therefore does not have strong affective or intuitive implications."[3] This research suggests that precise numerical likelihood and contextual information interact to affect judgment and decision making.

Several different factors will affect how a decisionmaker perceives a risk forecast. The first are characteristics of the risk forecast. Both the nature of the narrative evidence and the format of the likelihood assessment (e.g., verbal, percentage, or frequency representations of likelihood) may influence how decisionmakers feel about the quality of a forecast and the risk associated with the forecasted event. For instance, it is not clear whether decisionmakers will better use frequency (i.e., 1 out of 10) or standard likelihood representations (i.e., 1 or 10%) when evaluating likelihood assessments in simple forecasting contexts. Waters *et al*.[7] found that participants made more accurate assessments of total risk in a medical trade-off situation when likelihood was reported as a percentage as opposed to a frequency; however, performance was very similar between percentage and frequency formats across a range of operations.[1] In fact, recent research suggests that the usefulness of percentage versus frequency formats depends on the specific operations required of the decisionmaker.[12]

The skills and characteristics of the individual decisionmakers may also affect how a decisionmaker perceives a risk forecast. Although several individual difference factors and competencies may affect perceptions of risk forecasts (e.g., worldview, political stance, etc.), we focus on numeracy. Numeracy defined in the broadest sense is the ability to understand and use numbers. In this work, numeracy is defined as the ability to understand and manipulate proportions, risks, percentages, and probabilities.[13] Decisionmakers lower in numeracy have been shown to have trouble evaluating numerical information and integrating this information into their risk perceptions, health and financial decisions, and other real-world judgments and decisions.[14–17] However, how numeracy influences perceptions and evaluations of risk forecasts has not been studied.

How might numerical ability affect decisionmakers' understanding and use of risk forecasts? First, all decisionmakers may try to use the stated likelihood in a forecast but the less numerate may evaluate probabilities in a more imprecise and idiosyncratic way (i.e., less numerate decisionmakers may have very different evaluations of a 10% chance). In addition, because less numerate participants may have difficulty evaluating numerical probabilities, they may also be more affected by how the likelihood information is communicated. Peters *et al.*[18] examined how individuals varying in numeracy were able to understand and use probabilities expressed in different formats and to what extent they were affected by information framing. Less numerate decisionmakers were more affected by framing manipulations and interpreted identical likelihood information differently depending on whether it was in frequency (10 out of 100) or percentage form (10% out of 100).

Second, because less numerate participants have more difficulty evaluating probabilities, they may underweight the stated likelihood and focus on other nonnumeric information presented in the forecast or base their judgments on preconceptions. In the forecasting context, we may expect less numerate decisionmakers to use the narrative evidence in a forecast to a greater extent. Findings from Peters and colleagues suggest that the less numerate tend to be more influenced by irrelevant nonnumeric sources of information when making decisions.[18–20] This is important because the information that is used when evaluating a risk forecast has the potential to greatly influence risk perceptions and risk management decisions.

In addition, the extent to which a decisionmaker perceives a forecast to be useful and trustworthy will likely affect how much the forecast will be used. Thus, it is important to identify any differences in perceptions of usefulness and trust among decisionmakers varying in numeracy.

[1] Gigerenzer and colleagues have argued that natural frequency representations of probability are more amenable to clear understanding and improve statistical thinking in Bayesian inference problems.[8–10] However, natural frequency representations are distinct from normalized representations of probability (i.e., probability representations with base rate information removed as in 20% or 20 out of 100[11]). We are studying normalized single-event probability assessments that are typically used by experts in forecasts.

In two experiments, we explore how decisionmakers varying in numeracy use narrative evidence and stated likelihood in intelligence risk forecasts. Our results add to the growing numeracy literature by examining the mechanisms by which numeracy affects judgment and decisions in risk communication contexts. We also contribute to the growing literature on how people use likelihood in different formats. These results are also relevant to practicing forecasters who should be aware of the effects that individual differences in numeracy can have on the interpretations of their forecasts.

1. STUDY 1

In this study, we explored how decisionmakers varying in numeracy used explicit likelihood assessments with and without narrative evidence in a simulated terrorism risk forecast. In simple forecasts without narrative evidence, we expected the stated likelihood information to be particularly salient. Examining this forecasting context allows us to test hypotheses concerning imprecise evaluations of likelihood by the less numerate in a more straightforward manner without the narrative evidence affecting perceptions of the forecast.

HYPOTHESIS 1: To the extent that decisionmakers lower in numeracy evaluate probabilities in more imprecise and idiosyncratic ways we would expect greater variance in the perceived risk ratings of the less numerate.

HYPOTHESIS 2: As previous research suggests, less numerate participants may also be more influenced by likelihood format because their evaluations of likelihood are more labile and imprecise.

HYPOTHESIS 3: In addition, if the less numerate have more imprecise evaluations of likelihood they may also show less consistent mean risk ratings for higher and lower stated likelihood. In other words, the less numerate may be less likely to distinguish between a forecast with a 5% versus 20% stated likelihood.

In more complex forecasts, the narrative evidence is another relevant, and likely powerful, source of information for decisionmakers. To the extent that decisionmakers are focusing more on the narrative evidence than stated likelihood, we would expect a smaller impact of the stated likelihood on choices and risk perceptions. Because less numerate participants are uneasy with numerical information, they may underweight the stated likelihood and focus on the narrative evidence presented in the forecast.

HYPOTHESIS 4: Thus, less numerate participants may be less impacted by the stated likelihood when evaluating forecasts that include narrative evidence.

1.1. Method

1.1.1. Procedure and Materials

A community sample from Eugene, Oregon was recruited and participants were paid $10 for a one-hour session that included other unrelated tasks. Each participant was asked to read one simulated intelligence report about a potential terrorist attack in Washington, DC (see Fig. 1). The report was presented with or without a narrative description of the evidence concerning the potential attack as well as a likelihood assessment and a statement about the potential lives lost and property damage if the attack were to occur. The statement about the potential lives lost and property damage in the worst-case scenario was held constant across conditions. After reading the intelligence report, participants responded to a series of questions about what they read. Participants also completed an extended numeracy scale.[13]

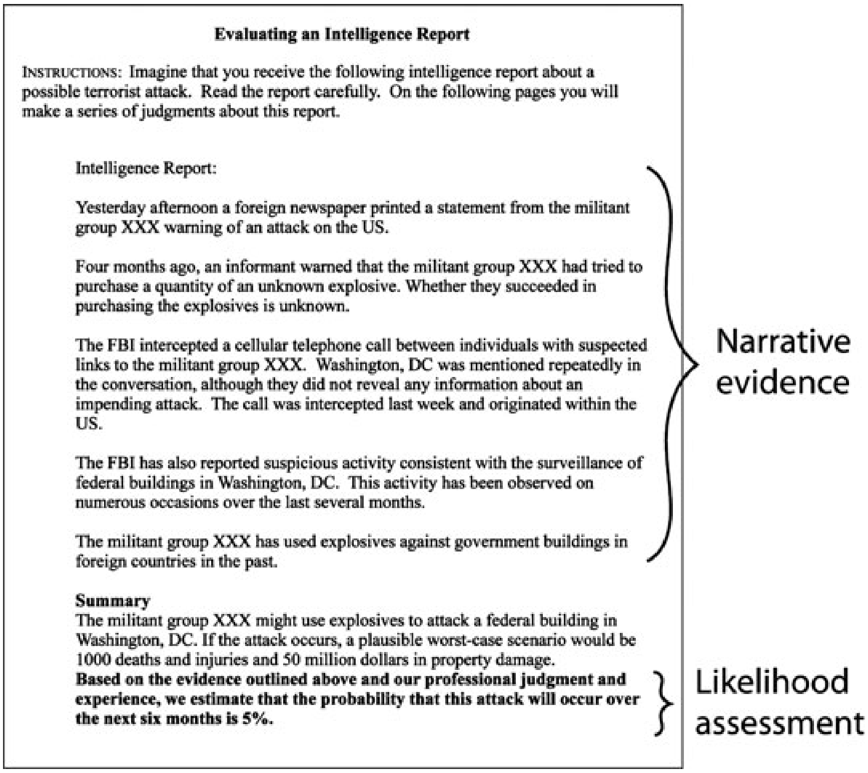
1.1.2. Experimental Design

The experiment was run fully between subjects in a 3 (uncertainty format) $\times$ 2 (stated likelihood) $\times$ 2 (narrative evidence) design. Stated likelihood was presented in a verbal, frequency, or percentage format, and likelihood was presented as either highly unlikely (5%, 5 out of 100) or fairly unlikely (20%, 20 out of 100).[2] Additionally, the intelligence report was presented with or without a narrative description of the evidence concerning the attack.

1.1.3. Dependent Variables

After reading the intelligence forecast, participants were asked to make the following choice:

[2] The verbal and numerical probability statements for each probability level were roughly matched based on previous research,[21,22] where highly unlikely was found to roughly correspond to a 5% probability and fairly unlikely was found to roughly correspond to a 20% probability. Bisantz *et al.*[23] have used a similar approach in matching verbal and numerical probability statements.



Evaluating an Intelligence Report

INSTRUCTIONS: Imagine that you receive the following intelligence report about a possible terrorist attack. Read the report carefully. On the following pages you will make a series of judgments about this report.

Intelligence Report:

Yesterday afternoon a foreign newspaper printed a statement from the militant group XXX warning of an attack on the US.

Four months ago, an informant warned that the militant group XXX had tried to purchase a quantity of an unknown explosive. Whether they succeeded in purchasing the explosives is unknown.

The FBI intercepted a cellular telephone call between individuals with suspected links to the militant group XXX. Washington, DC was mentioned repeatedly in the conversation, although they did not reveal any information about an impending attack. The call was intercepted last week and originated within the US.

The FBI has also reported suspicious activity consistent with the surveillance of federal buildings in Washington, DC. This activity has been observed on numerous occasions over the last several months.

The militant group XXX has used explosives against government buildings in foreign countries in the past.

Summary
The militant group XXX might use explosives to attack a federal building in Washington, DC. If the attack occurs, a plausible worst-case scenario would be 1000 deaths and injuries and 50 million dollars in property damage.
Based on the evidence outlined above and our professional judgment and experience, we estimate that the probability that this attack will occur over the next six months is 5%.

Fig. 1. The intelligence report used in Study 1. This report includes the narrative evidence and has a 5% assessed probability. The stated probability was varied at two levels (5% and 20%) and was presented in a verbal, frequency, or percentage form. The intelligence report was also presented with or without the narrative evidence.

> There are currently agencies providing security for federal buildings in Washington, DC. These agencies are using all of their available resources. The only way to increase security of the federal buildings would be to take resources from the local police department. When this was done in the past it resulted in an increase in violent crime in the city. Numerous violent crimes including several murders were blamed on the redistribution of the police forces. Given the intelligence assessment above, you need to make the following choice:
>
> 1. Leave the current security agencies in place.
>   OR
> 2. Take resources from the local police department to provide additional security.

Next, participants were asked to "rate the risk associated with this possible attack" on a 0–10 scale ranging from "very low risk" to "very high risk." Participants also rated the perceived usefulness of the report ("How valuable is this intelligence report? In other words, does it provide useful information for determining future actions to take?") and how knowledgeable they thought the analyst was ("How knowledgeable does this analyst seem about this potential attack?"). Participants rated perceived trust ("How much do you trust that this analyst is giving you complete and unbiased information/conclusions about this potential attack?"). All variables were rated on 0–10 scales, ranging from "not at all valuable/knowledgeable/trustworthy" to "extremely valuable/knowledgeable/trustworthy."

1.1.4. *Measurement of Numeracy*

A numeracy score was generated for each subject by summing the number of correct answers on an expanded numeracy scale (perfect score is 15). Numeracy was included as a continuous variable in the analyses, although a median split was sometimes used for simplicity of display.

1.1.5. *Analytic Approach*

The two primary dependent variables of choice and perceived risk were modeled with logistic regression and ANOVA models, respectively. A multivariate general linear model (MANOVA) was used to model the three perception variables of usefulness, knowledge, and trust because of the high correlations between these variables (rs = 0.56–0.71). Exploratory interaction effects were evaluated at $\alpha = 0.01$ to minimize the reporting of type-1 errors. Simple effect contrasts were also used to answer specific questions of interest. All of the results involving numeracy remained unchanged when controlling for

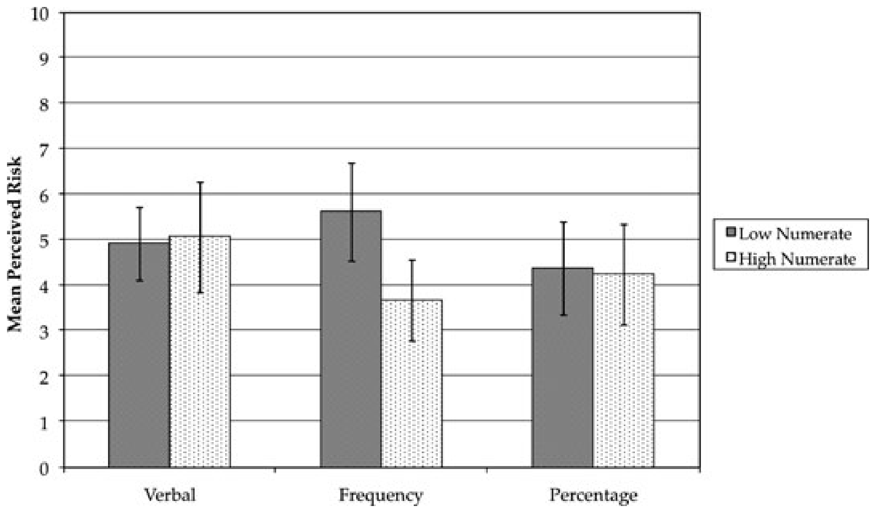


Fig. 2. Study 1: The effect of likelihood format and numeracy on risk perceptions in forecasts without narrative evidence.

education level and a proxy measure of verbal intelligence.

1.2. Results

1.2.1. *Sample Characteristics*

The mean age of the sample was 47.8 ($SD = 14.8$) and was 61% female ($n = 237$). Twenty-eight percent of the sample had a high school education or less, 40% had some college education, and 28% had a college degree or higher. The distribution of numeracy scores was negatively skewed (skewness = −0.54, $se = 0.16$; Mean = 9.66; $Md = 10.00$; Min = 2; Max = 15). Transforming the numeracy scores to reduce the negative skewness did not substantively change the results so the original scale was used.

1.2.2. *Forecasts Without Narrative Evidence*

To test the first set of hypotheses, we restrict our analysis to participants who saw the forecast without narrative evidence. Only the format and level of the stated likelihood were manipulated.

1.2.2.1. Explicit likelihood and likelihood format.After controlling for numeracy, participants who saw the higher stated likelihood in the forecast reported greater perceptions of risk, $F(1,107) = 8.221$, $p = 0.005$, $\eta^2 = .07$ (Higher likelihood mean = 5.27, Lower = 4.12), and were more likely to choose to divert forces from the local police, Odds ratio = 0.19, $p = 0.007$, 95% CI = (0.05, 0.64) (Higher percent = 28.5%, Lower = 12.3%). As a main effect, likelihood format did not significantly affect risk perceptions, choices, or perceptions of usefulness, knowledge, or trust.

1.2.2.2. Choices and perceptions by decisionmakers varying in numeracy. Decisionmakers lower in numeracy reported higher perceptions of risk, $F(1, 107) = 3.81$, $p = 0.05$, $\eta^2 = 0.034$ (More numerate mean = 4.28, Less = 4.95), and were more likely to choose to divert forces from the local police to protect against the potential attack, Odds ratio = 0.65, $p < .001$, 95% CI = (0.53, 0.79) (More numerate percent = 5.4%, Less = 35.1%). The less numerate ($SD = 2.046$) did not show significantly greater variance in risk ratings compared to the more numerate ($SD = 2.308$), $F(1, 117) = 3.324$, $p = 0.07$. Unexpectedly, it was the more numerate who showed a trend toward greater variance in risk ratings.

Because the less numerate may have more imprecise representations of likelihood, their risk perceptions may be more dependent on likelihood format than the more numerate.[15] Fig. 2 shows mean risk perceptions by likelihood format for the more and less numerate. The interaction between likelihood format and numeracy was marginally significant, $F(2,107) = 2.39$, $p = 0.097$, $\eta^2 = 0.043$, and simple effects were used to test for differences between the frequency and percentage formats based on numeracy. Previous research suggests that the more numerate will show more stable risk perceptions between alternative numerical likelihood formats than the less numerate. The less numerate do, in fact, show a larger difference in perceived risk between the frequency and the percentage conditions,

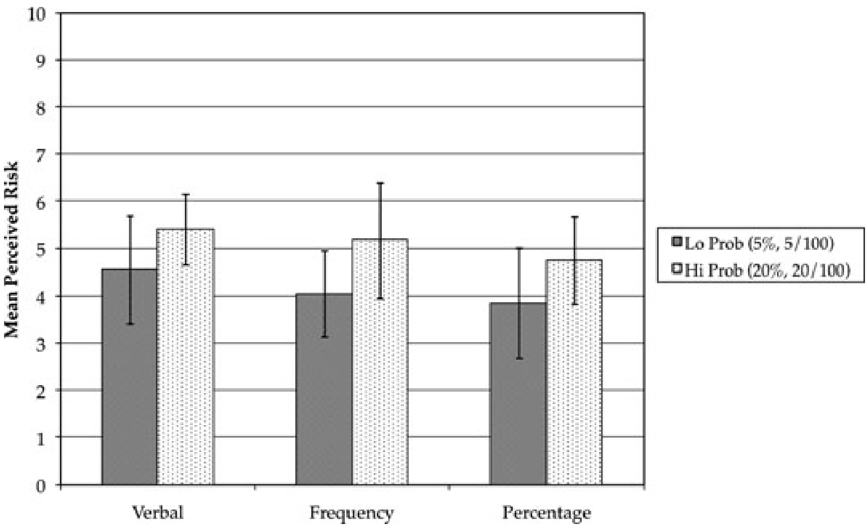


Fig. 3. Study 1: The effect of stated likelihood and likelihood format on perceived risk in forecasts without narrative evidence.

$F(1,107) = 3.33$, $p = 0.07$, $d = 0.58$, than the more numerate, $F(1,107) = 0.74$, $p = 0.40$, $d = 0.26$. In addition, more and less numerate decisionmakers only reported different risk perceptions when presented with the frequency format (see Fig. 2).

Next, we explored whether the more and less numerate were equally sensitive to differing levels of stated likelihood in the forecast (i.e., higher perceived risk in response to higher stated likelihood). Fig. 3 shows mean perceived risk by likelihood level for each likelihood format. The effect of likelihood level did not interact with likelihood format or numeracy. Thus, both the more and less numerate were able differentiate between different levels of likelihood in this simple probabilistic forecast without narrative evidence.

Numeracy level and numerical likelihood format interacted to affect perceptions of the quality and usefulness of forecasts, $F(3,105) = 3.14$, $p = 0.03$, Pillai's trace = 0.08.[3] The more numerate rated the forecast as more useful and the forecaster higher in knowledge and trust when presented with a percentage format (Mean usefulness = 5.24, knowledge = 4.48, trust = 4.38) as compared to a frequency format (Mean usefulness = 3.24, knowledge = 3.33, trust = 3.00). The less numerate rated the forecast as more useful and the forecaster higher in knowledge and trust when presented with the frequency format (Mean usefulness = 5.61, knowledge = 5.17, trust = 4.83) as compared to a percentage format (Mean usefulness = 4.74, knowledge = 4.79, trust = 3.47).

[3] Pillai's trace is the multivariate extension of univariate "percentage of variance accounted for" effect size measures (e.g., η^2).

1.2.3. Forecasts with Narrative Evidence

In this section, we restrict our analysis to participants who saw both the likelihood assessment and a set of narrative evidence accompanying the forecast.

1.2.3.1. Explicit likelihood and likelihood format. After controlling for numeracy, participants who saw the higher stated likelihood in the forecast with narrative evidence did not report greater perceptions of risk, $F(1,106) = 0.207$, $p = 0.65$, $\eta^2 = .002$ (Higher likelihood mean = 5.28, Lower = 5.08), although they were more likely to choose to divert forces from the local police, Odds ratio = 0.44, $p = 0.08$, 95% CI = (0.17, 1.09) (Higher likelihood percentage = 30.4%, Lower = 15.8%). As a main effect, likelihood format did not significantly affect risk perceptions, choices, or perceptions of usefulness, knowledge, or trust.

1.2.3.2. Choice, risk, and quality perceptions by decisionmakers varying in numeracy. Decisionmakers lower in numeracy reported higher perceptions of risk, $F(1, 106) = 3.42$, $p = 0.07$, $\eta^2 = 0.031$ (More numerate mean = 4.75, Less = 5.43), but they were not more likely to choose to divert forces from the local police to protect against the potential attack, Odds ratio = 0.95, $p = 0.54$, 95% CI = (0.81, 1.11) (More numerate percentage = 20.4%, Less = 25.0%). In contrast to the results without the narrative evidence, decisionmakers lower in numeracy (Mean usefulness = 5.93, knowledge = 5.27, trust = 4.61) reported that the forecast was more useful and the forecaster

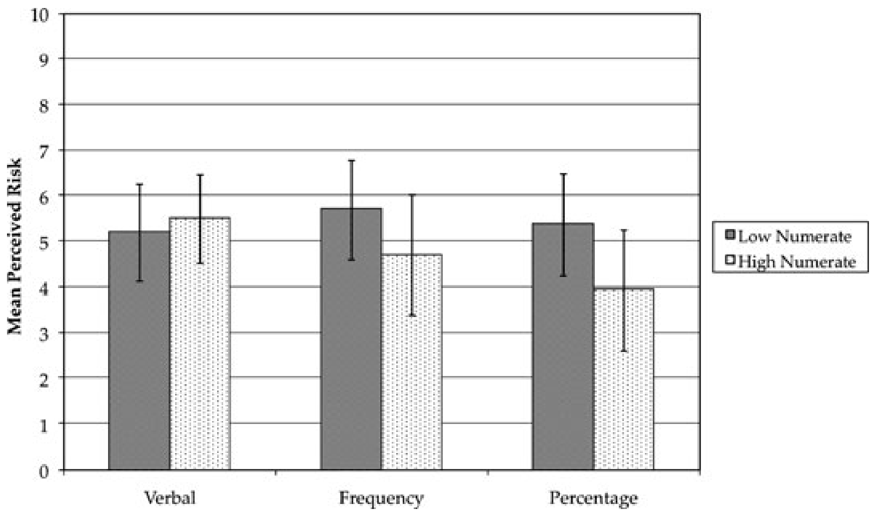


Fig. 4. Study 1: The effect of likelihood format and numeracy on risk perceptions in forecasts with narrative evidence.

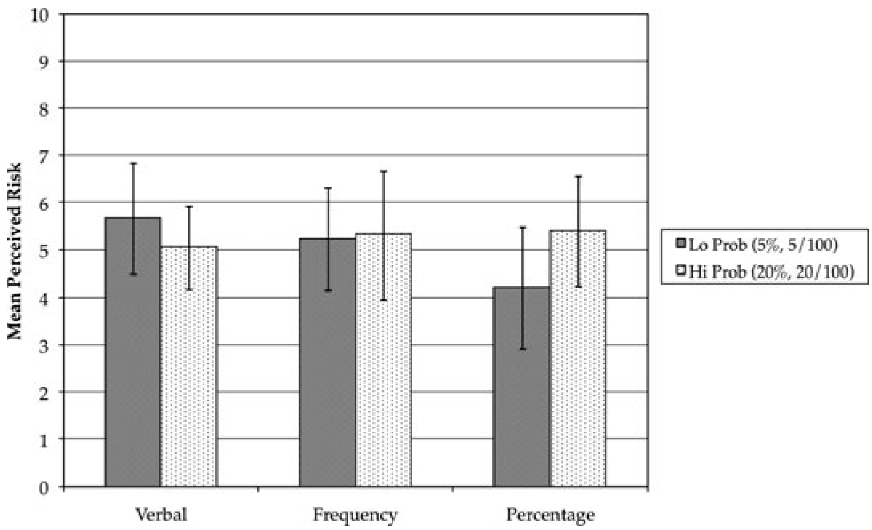


Fig. 5. Study 1: The effect of stated likelihood and likelihood format on perceived risk in forecasts with narrative evidence.

was more knowledgeable and trustworthy than did the more numerate (Mean usefulness = 4.90, knowledge = 4.29, trust = 3.76), $F(3,104) = 3.57$, $p = 0.016$, Pillai's trace = 0.09. Again, the less numerate (SD = 2.46) did not show significantly greater variance in risk ratings compared to the more numerate (SD = 2.38), $F(1, 116) = 0.39$, $p = 0.54$.

To the extent that participants focused on the narrative evidence as opposed to the stated likelihood we would expect a decrease in the effect of stated likelihood and likelihood format. Fig. 4 shows mean risk perceptions by likelihood format for the more and less numerate. Numeracy and likelihood format interacted to affect perceived risk, $F(2,106) = 4.70$, $p = 0.01$, $\eta^2 = 0.08$. The less numerate showed little difference in risk perceptions between the likelihood formats. The more numerate showed higher risk perceptions in the verbal likelihood condition compared to the numerical likelihood formats, verbal versus percentage, $F(1,106) = 3.93$, $p = 0.05$, $d = 0.71$, verbal versus frequency, $F(1,106) = 0.95$, $p = 0.33$, $d = 0.35$. This result suggests that they were more sensitive to the presence of numbers than the less numerate in the context of a narrative.

Fig. 5 shows mean risk perceptions for likelihood level and likelihood format. The inconsistent nature of the risk perceptions suggests that participants were not paying as much attention to the stated likelihood in the presence of the narrative evidence. In fact, mean risk perceptions were only consistent with the levels of stated likelihood in the percentage likelihood condition. These effects did not interact with numeracy, however, which may be due to insufficient power.

1.3. Discussion

In forecasts without narrative evidence we expected the explicit likelihood assessments to be particularly salient to decisionmakers. To the extent that decisionmakers lower in numeracy evaluate likelihood assessments in more imprecise and idiosyncratic ways, we expected the less numerate to show greater variance in risk ratings, be more influenced by the likelihood format, and be less consistent distinguishing between higher and lower stated likelihood levels.

In a simple forecast without narrative evidence, less numerate decisionmakers perceived greater risk and were more likely to divert forces from the local police. Less numerate decisionmakers did not, however, show greater variance in risk ratings. Consistent with previous research, the less numerate did appear to show less stable risk perceptions between different numerical likelihood formats. However, decisionmakers of all numeracy levels appeared to use the different levels of stated likelihood information to inform choices and risk perceptions. These results suggest that the less numerate have more labile evaluations of likelihood information in forecasts that depend on likelihood format. They also reported higher risk perceptions overall. However, the less numerate did show sensitivity to differing levels of stated likelihood. This indicates that in a simple forecasting situation like this one, any imprecision in the evaluation of stated likelihood by the less numerate did not substantially affect their ability to use this information.

We expected the narrative evidence to be a vivid and particularly salient piece of information for the decisionmakers to use when evaluating the forecast. If less numerate participants are more uncomfortable using explicit likelihood, they may focus on other easily evaluated information presented in the forecast. To the extent that participants are focusing on the narrative evidence, we may expect a smaller impact of the stated likelihood on choices and risk perceptions.

In this more complex forecast with narrative evidence, less numerate decisionmakers still perceived greater risk overall than the more numerate, but they were not more likely to choose to divert forces from the local police. Likelihood format did not affect the risk perceptions of the less numerate, suggesting perhaps that the presence of the narrative evidence decreased their attention to the likelihood information. Less numerate decisionmakers also reported greater usefulness, knowledge, and trust ratings than the more numerate. Since this effect was not present when this identical forecast did not have narrative evidence, this suggests that the less numerate thought that the addition of the narrative evidence was more valuable than the high numerate did. However, decisionmakers at all levels of numeracy did not use the stated likelihood as consistently with narrative evidence present. The results suggest that all decisionmakers were sensitive to the narrative evidence when it was included in the forecast and it appeared to decrease the use of stated likelihood. Although the evidence is inconclusive in this study, the less numerate may be more likely to use the narrative evidence. The next study was designed to more directly test the hypothesis of differential information use by decisionmakers varying in numeracy. In Study 2, we examined the relative weight that decisionmakers place on the narrative evidence and explicit likelihood and whether the relation between these evidence sources and likelihood ratings was moderated by numeracy.

These results also have some interesting implications for the use of different likelihood formats in the forecasting domain. Without the narrative evidence, decisionmakers were sensitive to the likelihood information in all formats. With narrative evidence present, however, decisionmakers appeared to be sensitive to only the percentage likelihood format and not the verbal or frequency formats. The fact that verbal likelihood statements were not used well is not surprising in light of previous research findings, although it is interesting that they were used consistently when presented without narrative evidence.[24–27] The apparent superiority of the percentage format over the frequency format does add to the conflicting literature on the usefulness of these formats. In particular, these results suggest that when additional narrative evidence in available to the judge, a percentage likelihood format will be used to a greater extent than frequency or verbal formats.

2. STUDY 2

Study 2 was designed to more directly assess decisionmakers' use of narrative evidence and stated likelihood when forecasts include both of these information sources. We were interested in perceived characteristics of the narrative evidence that may have a strong influence on risk perceptions. For instance, the coherence as well as the credibility of the evidence will likely affect how decisionmakers perceive the likelihood and overall severity of a forecasted event. The coherence of the evidence is how well the evidence set can be formed into a coherent, causal story that points toward the forecasted event occurring (i.e., the implication of the evidence). The credibility of the evidence is how much the individual pieces of evidence can be trusted for accuracy (i.e., the strength or weight of the evidence). In this study, we asked decisionmakers about their perceptions of the coherence and credibility of the evidence set to explore how these perceptions related to perceptions of likelihood, usefulness, and source credibility.

The primary goal of this experiment was to test whether decisionmakers with differing levels of numeracy focused on different sources of information when judging the likelihood of a forecasted event. Based on previous findings,[18,20] we hypothesized that decisionmakers lower in numeracy will focus more on their perceptions of the narrative evidence and those higher in numeracy will focus more on the explicit likelihood assessments.

2.1. Method

2.1.1. Procedure and Materials

Graduate students attending the University of Oregon were paid $14 for a one-hour experimental session. A more educated sample was used in this study to more closely match the demographic that would likely encounter risk forecasts in their work. Participants were presented with simulated intelligence forecasts that warned of a possible terrorist attack. Three new terrorism scenarios were generated (see Appendix B). The scenarios were very similar in terms of written length and the number and types of evidence used. Participants responded to a series of questions about each scenario and completed the same numeracy measure used in Study 1.

2.1.2. Experimental Design

In Study 1, explicit likelihood was only presented at two levels (5% and 20%). To further test the sensitivity of decisionmakers to explicit likelihood in risk forecasts and to allow more detailed analysis of the function relating explicit to perceived likelihood, three levels of likelihood (i.e., 1%, 5%, 10%) were used in this study. In addition, the results from Study 1 suggested that, in the presence of a narrative description of the evidence, the verbal and frequency formats would not elicit consistent ratings of risk. To maximize the potential use of stated likelihood only the percentage format was used in this study.

This experiment was run as a 2 (likelihood format) $\times$ 4 (likelihood level) mixed experimental design with likelihood level as the within-subject factor. The four levels of stated likelihood were: no stated likelihood, 1%, 5%, and 10%. The likelihood format factor varied as follows: (1) a point estimate of likelihood framed as an external estimate ("The probability that this event will occur is $\times$% ..."), (2) a point estimate of likelihood framed as a rating of how confident the analysts are that the event is going to occur ("We are $\times$% sure that this event will occur ..."). The pairing of scenario to likelihood level as well as the order of presentation was randomized to control for incidental effects. Since our goal in this study is to test differential information use by decisionmakers varying in numeracy, we do not discuss the condition without stated likelihood further in this article.

Each forecast had a set of narrative evidence, a statement about the numerical likelihood of the event, and a statement about the potential threat or harm that would result if the attack were to occur. The statement about potential harm was held constant for all of the reports: "If the attack occurs, a plausible worst-case scenario would be 1000 deaths and injuries and 50 million dollars in property damage."

2.1.3. Dependent Variables

Participants were asked to rate their perceived likelihood of the attack after reading each intelligence forecast (0–100%). The lower bound of the rating scale was labeled with "No chance," 50% was labeled with "As likely as unlikely," and the upper bound was labeled with "Certain." In addition, decisionmakers were asked about the perceived usefulness of the forecast for decision making. Finally, source credibility was assessed with a scale used by McComas and Trumbo.[28] The source credibility scale is made up of five questions asking the extent to which decisionmakers trust the conclusions of the forecast, whether they feel the forecast is accurate, whether it is fair, whether it tells the whole story, and whether it is biased.

After reading each of the four intelligence reports and responding to the primary measures described above, the decisionmakers were asked to go back through the materials and make two additional ratings about the evidence described in each scenario. The first was a global rating of the overall credibility of the evidence in each scenario, and the second was a rating of how well the evidence in each scenario could be formed into a coherent story. Participants responded to these questions on 11-point rating scales anchored by "very low" and "very high" coherence/credibility.

2.1.4. *Analytic Approach*

The data were fit with several multilevel mixed models, with the repeated measures represented at level 1 (i.e., likelihood level and ratings of the coherence and credibility of the evidence) and the between-subject data represented at level 2 (i.e., the likelihood format and numeracy). The multilevel modeling approach is particularly useful when dealing with covariates that vary within subjects, which can be difficult to model with standard general linear models.[29] All inferential tests were conducted with robust standard errors, which further guard against the influence of violating critical assumptions.[29] Exploratory interaction effects were evaluated at $\alpha = 0.01$ to minimize the reporting of type-1 errors.

The effect of likelihood level is represented as a linear slope (higher order polynomial effects were not significant), which represents the extent to which decisionmakers perceived greater likelihood as stated likelihood increased. Ratings of the credibility and coherence of the evidence were highly correlated with one another (rs = 0.58–0.78). On the basis of the similar pattern of relations between the variables and model parsimony, the credibility and coherence ratings were averaged to create a composite variable. The likelihood ratings were SQRT transformed in all analyses to reduce extreme positive skewness. Numeracy was significantly negatively skewed due to the high education level of participants (skewness = −1.59, se = 0.31; Mean = 12.29; Md = 13.00; SD = 2.12; Min = 5; Max = 15). Thus, numeracy was square transformed (i.e., numeracy^2). The effects of numeracy on perceptions of likelihood remained unchanged after controlling for self-reported knowledge about political and world events, self-reported political stance (i.e., liberal-conservative), and perceptions of usefulness and source credibility of the forecast. Education was not controlled because all participants were highly educated in this sample.

2.2. Results

2.2.1. *Sample Characteristics*

The sample ($n = 58$) was 53% female with an average age of 28 yrs (SD = 5.8). Participants all had four-year college degrees and a majority were current graduate students attending the University of Oregon (91%).

2.2.2. *The Impact of Explicit Likelihood and Narrative Evidence on Perceptions of Likelihood*

In the first model, likelihood ratings were modeled as a function of stated likelihood and likelihood format. Decisionmakers rated an attack as more likely the higher the stated likelihood assessment, slope = 0.46, $t(170) = 3.28$, $p = 0.002$. Running this model with the raw likelihood ratings resulted in an intercept of 23.41 and a slope of 2.38 ($p = 0.05$). Thus, for each step increase in stated likelihood (i.e., 1%–5%–10%), decisionmakers perceived a 2.38 unit increase in perceived likelihood on a 0–100 scale. There was no significant main effect of likelihood format, $t(56) = 0.22$, $p = 0.83$. Since there were no main or interaction effects involving likelihood format, it will not be discussed further.

In the next model, the coherence and credibility of the narrative evidence and individual differences in numeracy were added as further predictors of perceived likelihood. Decisionmakers who rated the evidence in a forecast as higher in coherence and credibility perceived the attack to be more likely, slope = 0.38, $t(153) = 4.21$, $p < 0.001$. For each unit increase in perceived credibility and coherence, decisionmakers perceived a 3.19 unit increase in perceived likelihood (raw likelihood slope = 3.19, $p < 0.001$). Participants lower in numeracy reported greater levels of perceived likelihood, slope = −0.02, $t(51) = -2.68$, $p = 0.01$. For each unit decrease in numeracy, decisionmakers perceived a 2.92 unit increase in perceived likelihood (raw likelihood slope = −2.92, $p = 0.05$). As hypothesized, the effect of stated likelihood was smaller for decisionmakers lower in numeracy, $t(153) = 3.53$, $p = 0.001$, and the relationship between the average of the credibility and coherence ratings and perceived likelihood was higher for decisionmakers lower in numeracy, $t(153) = -2.12$, $p = 0.035$. Figs. 6 and 7 show the effects of stated likelihood and perceived credibility and coherence on perceived likelihood for decisionmakers with different levels of numeracy. These results suggest that, as

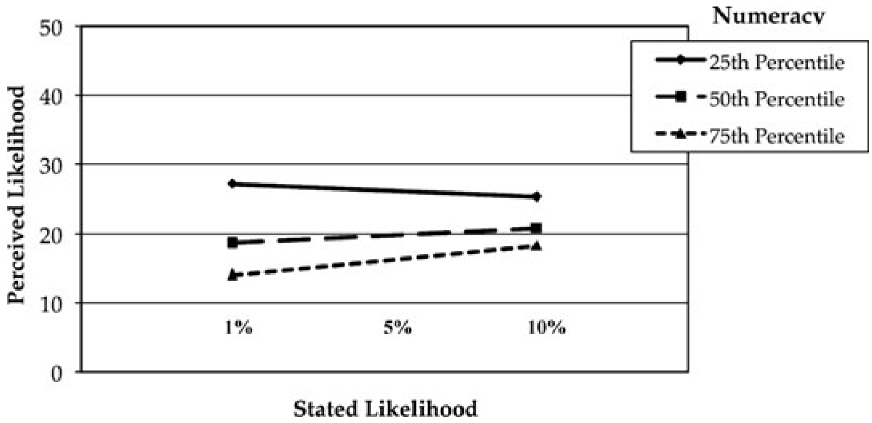


Fig. 6. Study 2: Model-based slopes showing the effect of stated likelihood on perceived likelihood for decisionmakers with different levels of numeracy.

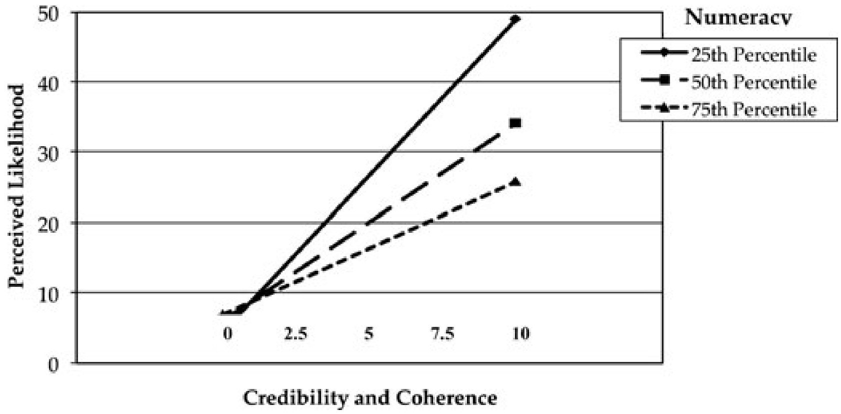


Fig. 7. Study 2: Model-based slopes showing the relation between perceived credibility/coherence and perceived likelihood for decisionmakers with different levels of numeracy.

hypothesized, decisionmakers lower in numeracy were more sensitive to the perceived properties of the narrative evidence and less sensitive to the stated likelihood information.

2.2.3. *Perceived Usefulness and Source Credibility*

Decisionmakers who rated the evidence set as more coherent and credible reported higher levels of perceived usefulness, $t(152) = 7.94$, $p < 0.001$, and source credibility, $t(153) = 7.54$, $p < 0.001$.

2.3. Discussion

As in Study 1, decisionmakers were sensitive to both the stated likelihood assessments as well as the narrative evidence. As stated likelihood increased, perceived likelihood of the attack increased as well. As perceptions of the credibility and coherence of the narrative evidence increased, so did perceived likelihood. Greater perceived credibility and coherence of the evidence was also related to greater perceived usefulness and source credibility. As in Study 1, decisionmakers lower in numeracy reported the forecasted attack to be more likely (higher risk in Study 1).

These results also provide evidence supporting our primary hypothesis that the less numerate would use their perceptions of the narrative evidence more when evaluating the terrorist risk whereas the more numerate would focus more on stated likelihood. These results are consistent with and extend previous research that shows differential information use between more and less numerate decisionmakers. However, since both the stated likelihood and narrative evidence are valid sources of information in the forecast, it is not clear whether it is best to focus on the stated likelihood or the narrative evidence. What is important is that decisionmakers with different levels of numeracy may focus on different information sources and potentially reach different conclusions regarding a forecasted event. To our knowledge, this is the first demonstration of differential information processing based on numeracy in a forecasting domain where both explicit likelihood and narrative evidence are available to the judge.

3. GENERAL DISCUSSION

Decisionmakers are often presented with explicit likelihood assessments and supporting narrative evidence in forecasting and risk communication domains. In most cases, decisionmakers likely rely on both numerical and narrative information to the extent that they perceive the information to be diagnostic, accurate, and trustworthy. In two studies, we explored how decisionmakers varying in numeracy evaluate likelihood statements in forecasts and if they focus on different information when both narrative evidence and stated likelihood information are available. We sampled community members (Study 1) and current graduate students (Study 2) covering a broad range of numerical ability.

The results from Study 1 suggested that less numerate decisionmakers do not show drastically more imprecise representations and were able to use numerical likelihood in simple forecasts without narrative evidence, although their perceptions of risk were more dependent on likelihood format. Decisionmakers varying in numeracy also perceived the forecast to be quite different in quality (i.e., usefulness, knowledge, and trust) depending on likelihood format. However, in Study 1 participants were asked about their perceptions of risk (not likelihood) and decisionmakers may have used other sources of information in addition to the stated likelihood. For instance, the worst-case consequence of the attack was presented in the forecast, and although it was the same at all levels of stated likelihood, idiosyncratic interpretations of this information may have increased the variance of risk ratings and decreased our ability to differentiate between the more and less numerate.

Across both studies, we observed the tendency for the less numerate to express higher risk and likelihood perceptions. This may be due to a more rich understanding by the more numerate of how small the probabilities in these studies actually were and the overall base rates for terrorist attacks in the population. In Study 2, we asked participants at the end of the experiment whether they had considered the base rate of terrorist attacks in the United States when judging the likelihood of specific terrorist plots they read about. Approximately 50% said that they had considered the base rate and those higher in numeracy were more likely to have done so, $r(52) = 0.30$, $p = 0.029$.

The less numerate also perceived the risk forecasts with narrative evidence to be higher in usefulness, knowledge, and trust than the more numerate, although this effect was not replicated in Study 2, perhaps due to the more educated sample and smaller range of numerical ability. The lower ratings of usefulness, knowledge, and trust may reflect the richer knowledge of the more numerate about the difficulty of forecasting events such as terrorist attacks in the real world and the particular difficulty in assigning likelihood point estimates to events that involve such deep uncertainty. This result also suggests that decisionmakers lower in numeracy may give undue weight to forecasts without acknowledging the inherent analytic uncertainty involved. This is one reason why analysts should report their level of analytic confidence in some fashion.[30]

In addition, we found that the less numerate appeared to use the narrative evidence to a larger extent than the more numerate, which could also explain their higher overall risk and likelihood ratings. The more numerate, on the other hand, used the explicit likelihood to a larger extent than the less numerate. Given the amount of numerical information that is presented to decisionmakers in this information-rich age, differences in numeracy have the potential to have a substantial effect on decisionmakers' perceptions and eventual decisions. These results are particularly important given that even highly educated individuals can be low in numeracy.[31] This has important implications for risk analysts and forecasters who need to report the results of their analyses to decisionmakers. Decisionmakers varying in numerical ability may arrive at very different perceptions of the risk of a forecasted event depending on what pieces of information they focus on. These different risk perceptions have the potential to lead to very different risk management decisions. Thus, even though an explicit likelihood estimate is presented, it does not mean that all decisionmakers are going to reach similar feelings about risk, particularly if they are focusing more on one source of relevant information over another. Additional research needs to be directed at how to present forecasts and risk analyses in a way that is interpretable to the majority of decisionmakers and accurately presents the opinions of the forecaster.

Firm conclusions from the present set of studies are limited by several factors. The stimuli used in both experiments were mock intelligence forecasts of potential terrorist plots involving explosive devices in the United States. Thus, it is possible that the results described are in some part restricted to hazards relating to terrorism presented in this specific form.

The sample of research participants may also limit the generalization of these results. Real consumers of these expert forecasts likely have specialized experience, knowledge, and backgrounds that may make them respond differently then lay consumers. The recruitment of more educated participants for Study 2 was done in part to simulate the likely education level of real intelligence consumers, although there are clearly other contextual factors that were not simulated in the current studies (e.g., worldview, political pressure that may affect a consumer's perception of a forecast, etc.). Another limitation to the generalizability of the results was the relatively narrow range of numerical likelihood that was manipulated (1–20%). The extent to which decisionmakers are sensitive to the numerical likelihood and narrative evidence may depend on the range of likelihood values presented. More research is needed on how consumers make judgments of likelihood when presented with a wider spectrum of assessed likelihood. Finally, the evidence supporting differential information use by decisionmakers varying in numeracy is correlational in nature. Future work should focus on manipulating properties of the supporting narrative evidence to test causal hypotheses and to identify the properties of the narrative evidence that have the largest effects on evaluations of forecasts.

ACKNOWLEDGMENTS

This research was conducted as part of Nathan Dieckmann's dissertation submitted to the University of Oregon. We would like to thank Robert Mauro and John Orbell who also served on the committee. We would also like to thank three anonymous reviewers for their helpful comments on previous versions of this article.

APPENDIX A: NUMERACY MEASURE

Numbers—you may not use a calculator for any of these questions.

1. Imagine that we roll a fair, six-sided die 1,000 times. Out of 1,000 rolls, how many times do you think the die would come up as an even number?

Answer:__________

2. In the BIG BUCKS LOTTERY, the chances of winning a $10.00 prize are 1%. What is your best guess about how many people would win a $10.00 prize if 1,000 people each buy a single ticket from BIG BUCKS?

Answer:__________ people

3. In the ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets of ACME PUBLISHING SWEEPSTAKES win a car?

Answer:__________%

4. Which of the following numbers represents the biggest risk of getting a disease?

_ 1 in 100 _ 1 in 1000 _ 1 in 10

5. Which of the following numbers represents the biggest risk of getting a disease?

_ 1% _ 10% _ 5%

6. If Person A's risk of getting a disease is 1% in 10 years, and Person B's risk is double that of A's, what is B's risk?

Answer:_____% in _____ years

7. If Person A's chance of getting a disease is 1 in 100 in 10 years, and person B's risk is double that of A, what is B's risk?

Answer:_____ in _____ years

8. If the chance of getting a disease is 10%, how many people would be expected to get the disease:

A: Out of 100? Answer:________ people

B: Out of 1000? Answer:________ people

9. If the chance of getting a disease is 20 out of 100, this would be the same as having a __% chance of getting the disease.

10. The chance of getting a viral infection is .0005. Out of 10,000 people, about how many of them are expected to get infected?

Answer:________ people

11. Which of the following numbers represents the biggest risk of getting a disease?

_ 1 chance in 12 _ 1 chance in 37

12. Suppose you have a close friend who has a lump in her breast and must have a mammography. Of 100 women like her, 10 of them actually have a malignant tumor and 90 of them do not. Of the 10 women who actually have a tumor, the mammography indicates correctly that 9 of them have a tumor and indicates incorrectly that 1 of them does not have a tumor. Of the 90 women who do not have a tumor, the mammography indicates correctly that 81 of them do not have a tumor and indicates incorrectly that 9 of them do have a tumor. The table below summarizes all of this information. Imagine that your friends tests positive (as if she had a tumor), what is the likelihood that she actually has a tumor?

	Tested Positive	Tested Negative	Totals
Actually has a tumor	9	1	10
Does not have a tumor	9	81	90
Totals	18	82	100

Answer: ________

13. Imagine that you are taking a class and your chances of being asked a question in class are 1% during the first week of class and double each week thereafter (i.e., you would have a 2% chance in Week 2, a 4% chance in Week 3, an 8% chance in Week 4). What is the probability that you will be asked a question in class during Week 7?

Answer: _______%

14. Suppose that 1 out of every 10,000 doctors in a certain region is infected with the SARS virus; in the same region, 20 out of every 100 people in a particular at-risk population also are infected with the virus. A test for the virus gives a positive result in 99% of those who are infected and in 1% of those who are not infected. A randomly selected doctor and a randomly selected person in the at-risk population in this region both test positive for the disease. Who is more likely to actually have the disease?

__ They both tested positive for SARS and therefore are equally likely to have the disease

__ They both tested positive for SARS and the doctor is more likely to have the disease

__ They both tested positive for SARS and the person in the at-risk population is more likely to have the disease.

APPENDIX B: THE FOUR INTELLIGENCE SCENARIOS USED IN STUDY 2

Intelligence Report 1

INSTRUCTIONS: Imagine that you receive the following intelligence report about a possible terrorist attack. Read the report carefully. On the following pages you will make a series of judgments about this report.

Intelligence Report:

Yesterday afternoon a foreign newspaper printed a statement from the militant group XXX warning of an attack on the United States.

Four months ago, an informant warned that the militant group XXX had tried to purchase a quantity of an unknown explosive. Whether they succeeded in purchasing the explosives is unknown.

The FBI intercepted a cellular telephone call between individuals with suspected links to the militant group XXX . Washington, D.C. was mentioned repeatedly in the conversation, although they did not reveal any information about an impending attack. The call was intercepted last week and originated within the United States.

The FBI has also reported suspicious activity consistent with the surveillance of federal buildings in Washington, D.C. This activity has been observed on numerous occasions over the last several months.

The militant group XXX has used explosives against government buildings in foreign countries in the past.

Summary

The militant group XXX might use explosives to attack a federal building in Washington, D.C.

Intelligence Report 2

INSTRUCTIONS: Imagine that you receive the following intelligence report about a possible terrorist attack. Read the report carefully. On the following pages you will make a series of judgments about this report.

Intelligence Report:

A tip from an anonymous informant recently led the FBI to the apartment of two men suspected of working for the militant group YYY. When the FBI arrived the men had already left, but investigators did discover simple maps and timetables of the railway systems in Chicago, IL.

Several months ago, a videotaped statement by the leader of the militant group YYY appeared on the Internet. Among other things, the leader alluded to a recent train bombing in Portugal and warned that the United States would be next.

Three months ago, analysts doing routine satellite monitoring of a known YYY training camp reported an increase in activity. It appeared that members of YYY were experimenting with explosive devices.

The YYY militant group has been implicated in several train bombings over the last several years. The most recent attack in Portugal was powerful enough to completely destroy one train car filled with passengers and completely derail the train.

Both the FBI and the Chicago Police have reported suspicious activity around train stops in the city. This activity has been observed on numerous occasions over the last several months.

Summary

The militant group YYY might use explosives to attack a train in Chicago. If the attack occurs, a plausible worst-case scenario would be 1,000 deaths and injuries and 50 million dollars in property damage.

Based on the evidence outlined above and our professional judgment and experience, we estimate that the probability that this attack will occur over the next six months is 1%

Intelligence Report 3

INSTRUCTIONS: Imagine that you receive the following intelligence report about a possible terrorist attack. Read the report carefully. On the following pages you will make a series of judgments about this report.

Intelligence Report:

On a few different occasions port authorities have stopped and questioned pairs of men trespassing in New York City ports. Each time the men were in the areas of the port where passenger ships dock.

A few weeks ago, a website with ties to the militant group ZZZ posted a statement that warned of attacks on the U.S. It specifically mentioned that the next attacks would be aimed at a vulnerable place, since so much security has been focused on air travel.

On a tip from an undercover agent, the FBI recently captured a wanted member of the militant group ZZZ. He revealed that group leaders had discussed attacking a port in New York City. He claimed to not know of any details concerning an attack and seemed unsure that members of the group had acquired the necessary explosives.

The FBI has also bugged the apartment of two suspected members of ZZZ. The men have been overheard discussing the technical details of previous terrorist attacks, as well as discussing preparations for leaving the city in the near future.

The ZZZ militant group has used explosives to attack targets in the past.

Summary

The militant group ZZZ might use explosives to attack a passenger ship in New York City. If the attack occurs, a plausible worst-case scenario would be 1,000 deaths and injuries and 50 million dollars in property damage.

Based on the evidence outlined above and our professional judgment and experience, we estimate that the probability that this attack will occur over the next six months is 5%.

Intelligence Report 4

INSTRUCTIONS: Imagine that you receive the following intelligence report about a possible terrorist attack. Read the report carefully. On the following pages you will make a series of judgments about this report.

Intelligence Report:

Several national security experts have predicted a terrorist attack during a large sporting event in the U.S. The high concentration of people in a relatively small area is the obvious draw of this type of attack.

A member of the militant group VVV was recently apprehended abroad. He revealed that the leadership of VVV had discussed several different plans to use explosives in the US. One plan was to coordinate several simultaneous explosive attacks in a highly populated area. Members of VVV have carried out attacks of this nature before.

In the last several months, both local authorities and the FBI have increased surveillance of professional basketball, baseball, football, and hockey events in the Los Angeles area. On one occasion, a suspicious package was left in a crowded area at a professional basketball game. The package turned out to be a hoax, but several authorities reported suspicious persons possibly observing the response. There is no way to be sure, but the hoax package could have been used to test the response of security and law enforcement.

Last week, the FBI confiscated financial statements and froze the bank accounts of a Los Angeles lawyer suspected of partially supporting members of VVV in the U.S. In the financial statements were records of a recent purchase of "mliitary materials." It is unknown what exactly was purchased or where the materials are located now.

Summary

The militant group VVV might use explosives to attack a professional sporting event in Los Angeles. If the attack occurs, a plausible worst-case scenario would be 1,000 deaths and injuries and 50 million dollars in property damage.

Based on the evidence outlined above and our professional judgment and experience, we estimate that the probability that this attack will occur over the next six months is 10%

Note: The examples above are the four different scenarios at each level of probability. In the experiment the pairing of probability level to each scenario as well as the order of presentation were randomized to control for incidental effects.

REFERENCES

1. Windschitl PD, Weber EU. 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, 1999; 25:1514–1533.
2. Windschitl PD, Martin R, Flugstad AR. Context and the interpretation of likelihood information: The role of intergroup comparisons on perceived vulnerability. Journal of Personality and Social Psychology, 2002; 82:742–755.
3. Flugstad AR, Windschitl PD. The influence of reasons on interpretations of probability forecasts. Journal of Behavior and Decision Making, 2003; 16:107–126.
4. Hendrickx L, Vlek C, Oppewal H. Relative importance of scenario information and frequency information in the judgment of risk. Acta Psychology, 1989; 72:41–63.
5. Hendrickx L, Vlek C, Calje H. Effects of frequency and scenario information on the evaluation of large-scale risks. Organizational Behavior and Human Decision, 1992; 52:256–275.
6. Hsee CK. The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alternatives. Organizational Behavior and Human Decision, 1996; 67:247–257.
7. Waters EA, Weinstein N, Colditz GA, Emmons K. Formats for improving risk communication in medical tradeoff decisions. Journal of Health Communication, 2006; 11:167–182.
8. Hoffrage U, Lindsey S, Hertwig R, Gigerenzer G. Communicating statistical information. Science, 2000; 290:2261–2262.
9. Gigerenzer G. Why the distinction between single-event probabilities and frequencies is relevant for psychology and (vice versa) Pp. 129–161 in Wright G, Ayton P (eds). Subjective Probability. New York: John Wiley & Sons, 1994.
10. Reyna VFB, Brainerd CJ. Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. Learning and Individual Differences, 2008; 18:89–107.
11. Hoffrage U, Gigerenzer G, Krauss S, Martignon L. Representation facilitates reasoning: What natural frequencies are and what they are not. Cognition, 2002; 84:343–352.

12. Cuite CL, Wedinstein ND, Emmons K, Colditz G. A test of numeric formats for communicating risk probabilities. Medical Decision Making, 2008; 28(3):377–384.
13. Peters E, Dieckmann NF, Dixon A, Hibbard JH, Mertz CK. Less is more in presenting quality information to consumers. Medical Care Research and Review, 2007; 64:169–190.
14. Nelson W, Reyna VF, Fagerlin A, Lipkus I, Peters E. Clinical implications of numeracy: Theory and practice. Annals of Behavior Medicine, 2008; 35:261–274.
15. Peters E, Hibbard JH, Slovic P, Dieckmann NF. Numeracy skill and the communication, comprehension, and use of risk and benefit information. Health Affairs, 2007; 26:741–748.
16. Peters E. Numeracy and the perception and communication of risk. Pp. 1–7 in Tucker WT, Ferson S, Finkel AM, Slavin D (eds). Annals of the New York Academy of Sciences. Strategies for Risk Communication: Evolution, Evidence, Experience. Vol 1128. New York: New York Academy of Sciences, 2008.
17. Reyna VF, Nelson W, Han P, Dieckmann NF. How numeracy influences risk reduction and medical decision making. Psychological Bulletin, under review.
18. Peters E, Västfjäll D, Slovic P, Mertz CK, Mazzocco K, Dickert S. Numeracy and decision making. Psychological Science, 2006; 17:407–413.
19. Peters E, Levin IP. Dissecting the risky-choice framing effect: Numeracy as an individual-difference factor in weighting risky and riskless options. Judgment and Decision Making, 2008; 3:435–448.
20. Peters E, Dieckmann NF, Västfjäll D, Mertz CK, Slovic P, Hibbard JH. Bringing meaning to numbers: The impact of evaluative categories on decisions. Journal of Experimental Psychology: Applied, in press.
21. Kent S. Words of estimated probability. In Steury DP (ed). Sherman Kent and the Board of National Estimates: Collected Essays. Center for the Study of Intelligence, 1994.
22. Hamm RM. Selection of verbal probabilities: A solution for some problems of verbal probability expression. Organizational Behavior and Human Decision, 1991; 48:193–223.
23. Bisantz AM, Marsiglio SS, Munch J. Displaying uncertainty: Investigating the effects of display format and specificity. Human Factors, 2005; 47(4):777–796.
24. Wallsten TS, Bedescu DV, Zwick R. Comparing the calibration and coherence of numerical and verbal probability judgments. Management Science, 1993; 39:176–190.
25. Fischhoff B. Learning from experience: Coping with hindsight bias and ambiguity. Pp. 543–554 in Armstrong JS (ed). Principles of Forecasting: A Handbook for Researchers and Practitioners. Boston: Kluwer Academic, 2001.
26. Heuer RJ. Psychology of Intelligence Analysis. Langley, VA: Central Intelligence Agency, Center for the Study of Intelligence Analysis, 1999.
27. Fox CR, Irwin JR. The role of context in the communication of uncertain beliefs. Basic and Applied Social Psychology, 1998; 20:57–70.
28. McComas KA, Trumbo CW. Source credibility in environmental health-risk controversies: Application of Meyer's credibility index. Annals of Behavioral Medicine, 2001; 35:261–274.
29. Raudenbush SW, Bryk AS. Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd ed. Thousand Oaks, CA: Sage, 2002.
30. Dieckmann NF, Mauro R, Slovic P. The effects of presenting imprecise probabilities in terrorism forecasting. Risk Analysis, under review.
31. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Medical Decision Making, 2001; 21:37–44.