Journal of Experimental Psychology: Applied
2005, Vol. 11, No. 4, 277–288

DOI: 10.1037/1076-898X.11.4.277

Are Risk Assessments of a Terrorist Attack Coherent?

David R. Mandel
Defence Research and Development Canada and University of Toronto

Four experiments examined 3 types of violations of coherence criteria in risk assessments of a terrorist attack. First, the requirement that extensionally equivalent descriptions be assigned the same probability (i.e., additivity) was violated. Unpacking descriptions of an attack into subtypes led to an increase in assessed risk. Second, additivity was also violated when risk assessments were obtained by subtracting the probability of no attack from 1.0. This refocusing procedure inflated assessed risk. Third, refocusing also increased the proportion of monotonicity violations in assessing risk across increasing or decreasing timeframes. Task structuring that promoted consideration of complementary possibilities increased coherence, suggesting that incoherence is due primarily to errors in applying rather than comprehending the relevant criteria.

Keywords: risk assessment, coherence violations, additivity property, monotonicity, terrorism

Since the September 11, 2001, terrorist attacks in the United States (U.S.) and the subsequent attacks in Bali, Madrid, London, Egypt, and elsewhere, many individuals have become acutely aware of the threat posed by terrorism (Huddy, Feldman, Capelos, & Provost, 2002). The economic, sociopolitical, and psychological consequences of this chronic focus on impending doom and its preemption are highly significant, ongoing, and tied to both the public's and public officials' threat perceptions and assessments of future risks of attack. How do members of the public assess risks such as the possibility of a major terrorist attack? Past research indicates that the magnitude of perceived risk increases with media exposure, feelings of dread, expectations of fatality, and a perceived lack of control (Slovic, 1987; for reviews, see Fischhoff, Bostrom, & Quadrel, 2002; Slovic, 2000). In the recent milieu of terrorist threat, these factors are exacerbated. Adding to a collective sense of uncontrollability, federal agencies, such as the FBI and CIA in the U.S., acknowledge the difficulties inherent in accurately predicting when, where, or how an attack might occur. Moreover, a considerable amount of news coverage since 9/11 has highlighted our vulnerabilities to terrorist attacks. Some of this coverage may trigger fear, which tends to increase perceived risk, while other coverage may provoke anger, which tends to decrease perceived risk (Lerner, Gonzalez, Small, & Fischhoff, 2003). Past research also indicates that the perception and assessment of risk can be perturbed by the manner in which such thoughts about risk are elicited (Fischhoff & MacGregor, 1983) and by the manner in which the judged events are described or framed (Slovic, Monahan, & MacGregor, 2000; Stone, Yates, & Parker, 1994).

The present article examines the coherence of risk assessments of a major terrorist attack in the U.S. Coherence in the present context refers to the internal consistency of risk assessments. Coherence may be contrasted with accuracy or correspondence (Hammond, 2000), which in the present context refers to the external consistency of risk assessments. The study of coherence violations in judgment and decision making research is most clearly associated with the heuristics and biases program of research initiated by Kahneman and Tversky (for overviews, see Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982). This body of research has often been ascribed the aim of demonstrating that individuals violate fundamental principles of logic that underscore normative theories of judgment and decision making.

Although studies of coherence have sometimes been described as having little to do with judging and deciding in the "real world" (e.g., Gigerenzer, 1991; Hammond, 2000), this body of research is in fact of considerable relevance to applied domains. Whereas accuracy may be viewed as the ultimate criterion of risk assessment, gauging it is not always feasible. All too often, the true probability of an event is unknown and may even be unknowable. The assessment of asymmetric threats, such as gauging the probability of a future terrorist attack in the U.S., is a case in point. Risk assessors may lack reliable statistical evidence and viable causal models needed to formulate accurate risk assessments. Influence models can systematize existing knowledge (e.g., Paté-Cornell & Guikema, 2002), but the predictive accuracy of these models

David R. Mandel, Judgment and Decision Making Group, Command Effectiveness and Behaviour Section, Defence Research and Development Canada, Toronto, Ontario, Canada, and Department of Psychology, University of Toronto, Toronto, Ontario, Canada.

Portions of the present research were presented in November 2004 at the Annual Meeting of the Society for Judgment and Decision Making in Minneapolis, Minnesota; in April 2005 at Defence Research and Development Canada's Science and Technology Symposium on Disruptive Technologies for the New Military in Ottawa, Ontario, Canada; in April 2005 at the Southern Ontario Behavioral Decision Research Conference in Waterloo, Ontario, Canada; and in July 2005 at the Twenty-Eighth Annual Scientific Meeting of the International Society for Political Psychology in Toronto, Ontario, Canada. This research was funded in part by Discovery Grant 249537-02 from the Natural Sciences and Engineering Council of Canada. I thank Phillip Ackerman, Peter Ayton, and Bobbie Spellman for their feedback on earlier drafts of this article.

Correspondence concerning this article should be addressed to David R. Mandel, Judgment and Decision Making Group, Command Effectiveness and Behaviour Section, DRDC Toronto, 1133 Sheppard Avenue West, P.O. Box, 2000, Toronto, Ontario M3M 3B9, Canada. E-mail: david.mandel@drdc-rddc.gc.ca

depends on the reliability and validity of the estimates plugged into the model and on the manner in which that information is integrated. When accuracy cannot be readily gauged, coherence may be used as a performance benchmark for risk assessors. Indeed, an examination of coherence can provide some indication, albeit imprecise, of the prospect for accuracy because incoherent assessments imply that at least one of the assessments provided was inaccurate. Thus, the study of coherence violations in risk assessment is of both basic and applied concern.

Past studies have shown that the coherence of risk perception can be perturbed by the manner in which risks are described or "framed." For example, people are willing to pay more to reduce the probability of an improbable risk "in half" rather than reduce it from, say, .00002 to .00001 (Stone et al., 1994). People are also more likely to attend to the risk if it is described as occurring 10 times per 1,000 rather than 1 time per 100 (Slovic et al., 2000). When risk assessments are solicited—as they are in the present research—they are also likely to depend on the manner in which the relevant events are described and how questions are worded. These features—inconsequential according to the normative principles of descriptive and procedural invariance (Kahneman & Tversky, 1984)—are nevertheless likely to influence risk assessments. For instance, Fischhoff and MacGregor (1983) found that although there was strong ordinal consistency across different methods for eliciting health risk assessments (e.g., "How many people die out of each 100,000 who contract influenza?" vs. "For each person who dies of influenza, how many have it and survive?"), absolute assessments varied by as much as two orders of magnitude. In the present article, I examine three ways in which the coherence of risk assessments of a terrorist attack may be systematically biased by varying the description of events or the wording of questions: (a) additivity violations due to unpacking, (b) additivity violations due to refocusing, and (c) and monotonicity violations due to refocusing.

Additivity Violations Due to Unpacking

A fundamental principle in probability theory—and hence a key coherence principle in risk assessment—is the additivity property, which states that the sum of the probabilities assigned to the occurrences of n mutually exclusive and exhaustive events must equal the probability that one of the events will occur (Edwards, 1982). When the events in question are binary complements, such as "terrorist attack" and "no terrorist attack," the additivity property is given a straightforward objectivist interpretation: these events must sum to unity (i.e., 1.0 on the [0, 1] probability scale).

According to Tversky and colleagues' *support theory* (e.g., Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994), the additivity property can easily be violated because individuals assign probabilities to *descriptions* of events (called *hypotheses* in the theory) rather than to events themselves. Studies have shown that the degree to which a hypothesis is descriptively compressed—that is, the degree to which it is "packed" or "unpacked"—will affect its assessed probability such that unpacking tends to inflate assessments (e.g., Ayton, 1997; Fischhoff, Slovic, & Lichtenstein, 1978; Fox, Rogers, & Tversky, 1996; Koehler, Brenner, & Tversky, 1997; Tversky & Koehler, 1994). Specifically, the theory predicts two types of unpacking effects. An implicit unpacking effect occurs when unpacking a hypothesis (e.g., "a terrorist attack") into an explicit disjunction (e.g., "a terrorist attack by al-Qaeda or non-al Qaeda operatives") leads to an inflation of subjective probability. An explicit unpacking effect occurs when unpacking a hypothesis into an extensionally equivalent set of elementary hypotheses (e.g., "a terrorist attack by al-Qaeda" or "a terrorist attack by non-al Qaeda operatives") leads to an inflation of subjective probability. In both cases, unpacking yields *subadditive* probability assessments because the probability assigned to the implicit hypothesis is less than that assigned to the sum of its parts when it is unpacked.

This article extends research on unpacking by examining its effect on risk assessments in an important applied domain. In the context of terrorist threat, media attention regarding the risk of a future terrorist attack in the U.S. has largely focused on al Qaeda as the probable source of future attacks. In Experiments 1a and 1b, different groups of participants were asked to assess the probability of: (a) at least one major terrorist attack in the U.S. (hereinafter, attack), (b) at least one major terrorist attack in the U.S. that is plotted by al Qaeda (hereafter, attack-aq), or (c) at least one major terrorist attack in the U.S. that is not plotted by al Qaeda (hereinafter, attack-no-aq). For sake of brevity, I refer to the mean probability of attack as the mean *packed* assessment and the sum of the mean attack-aq and attack-no-aq assessments as the mean *unpacked* assessment. I predicted that unpacking the risk of attack into the explicit scenarios of attack by al Qaeda and attack by operatives other than al Qaeda would produce subadditivity in risk assessments. That is, unpacked assessments would, on average, be significantly greater than packed assessment, indicating a violation of coherence in risk assessments across descriptions. Such a violation of additivity would indicate that risk assessors need to be cognizant of the fact that their assessments may be influenced simply by the degree of specificity with which a judged threat is described or brought to mind.

Additivity Violations Due to Refocusing

In the present research, I also examined another description-related effect on risk assessment called the *refocusing effect*. The refocusing effect refers to the tendency to assign different probabilities to an event depending on whether the event is judged directly or indirectly via an assessment of its complement. For example, if an individual assesses the probability of no attack (Y) to be .80 in the next 2 months, then on the basis of the additivity property this individual's assessment of the probability of attack (X) should be .20. If the assessment differs—namely, $P(X) \neq 1 - P(Y)$—then the additivity property would be violated and a refocusing effect would be observed. Unlike the unpacking effect, the refocusing effect is inconsistent with support theory, which predicts that probability assessments of binary complements are additive. Although the prediction of additivity fit early empirical findings (e.g., Wallsten, Budescu, & Zwick, 1993), some recent studies have shown that probability assessments of binary complements are superadditive (e.g., Idson, Krantz, Osherson, & Bonini, 2001; Macchi, Osherson, & Krantz, 1999; Sloman, Rottenstreich, Wisniewski, Hadjichristidis, & Fox, 2004; Villejoubert & Mandel, 2002; Windschitl, Kruger, & Simms, 2003, Experiment 4). That is, the probability of X and Y sum to less than unity. These findings may be described as refocusing effects in which $P(X) < 1 - P(Y)$.

Although the assessed risk of a major attack in the U.S. has been heightened since 9/11, it is still likely to be judged much less probable than its "no attack" complement. Extensional logic would thus require the complementary event (i.e., "not being attacked") to be judged as relatively probable. However, in the current sociopolitical milieu in which people are keenly aware of the prospect of an attack, it may seem more reasonable to judge the complementary prospect of no attack as being moderately rather than highly probable. Thus, an individual who assesses the risk of at least one attack to be .20 might nevertheless assess the probability of no attack to be "only" about .60. This pattern of risk assessment would result in a superadditive refocusing effect.

To test this hypothesis, a fourth group of participants in Experiments 1a and 1b was asked to assess the probability that no major terrorist attack would occur in the U.S. (hereinafter, no-attack). I predicted that, on average, refocused assessments of the risk of attack derived by subtracting assessments of the probability of no-attack from unity would be significantly greater than packed assessments. Experiments 2 and 3 further examined whether superadditive refocusing represents, as Kahneman and Tversky (1982) put it, an error of comprehension or an error of application. If the refocusing effect represents a comprehension error (i.e., risk assessors do not understand the additivity property), then it should persist even if risk assessments of binary complements are made in near succession—a condition that would increase the transparency of the complementary relation between alternative hypotheses. By contrast, if the refocusing effect represents an application error, then one would expect it to be attenuated when the complementary relation between judged events is transparent.

Between-subjects designs such as that used in Experiments 1a and 1b naturally obscure the complementary relation by providing participants with only one hypothesis from a pair. Thus, the comprehension-application-error distinction may be better assessed using within-subjects designs in which the salience of the additivity principle could be manipulated through varying the transparency of the complementary relation. Past research indicates that considering counterexamples can be an effective tool for debiasing a number of judgment biases including inferences about population proportions (Olson, 1976, see Experiment 2, Group 4). overconfidence (Hoch, 1985; Koriat, Lichtenstein, & Fischhoff, 1980), confirmation bias in hypothesis testing (Wason, 1960), and explanation bias in which the mere act of explaining why a future event might occur increases its perceived probability of occurrence (Lord, Lepper, & Preston, 1984). Extrapolating from these studies, one might expect to replicate the refocusing effect within-subjects when the complementarity relation was obscured, but to find a (debiasing) attenuation of the effect when the complementarity relation was salient. I tested this hypothesis in Experiments 2 and 3. In Experiment 2 (low transparency), each hypothesis (viz., attack and no-attack) was embedded in a set of unrelated probability assessment tasks and these sets were separated by a distinct intervening task. Thus, I anticipated that Experiment 2 would replicate the refocusing effect observed in the earlier experiments. By contrast, in Experiment 3 (high transparency), the complementary hypotheses were assessed in close succession with no intervening task. Consistent with the application-error interpretation, and past research indicating the effectiveness of considering the opposite as a debiasing strategy, I predicted that the refocusing effect would be attenuated, if not eliminated, in Experiment 3.

Monotonicity Violations Due to Refocusing

The present research also examined how coherently individuals adjust their risk assessments over increasing or decreasing timeframes. In Experiments 1a, 1b, and 2, participants were asked to assess the relevant risk (e.g., attack) for timeframes of 2, 4, and 6 months. Extensional logic requires that risk assessments across timeframes be adjusted monotonically—namely, they should never decrease as timeframe increases, nor should they increase as timeframe decreases. For instance, an individual who assesses the risk of attack in 2 months to be .10 and the risk of attack in 4 months to be .05 would violate the monotonicity requirement and would violate extensional logic because the latter hypothesis subsumes the former as a proper subset.

A study of the perceived consequences of risky behaviors in adults and adolescents by Beyth-Marom, Austin, Fischhoff, Palmgren, and Jacobs-Quadrel (1993) revealed that, whereas participants articulated the consequences for engaging in a focal behavior in an affirmative manner, they were significantly more likely to articulate the consequences of not engaging in the behavior in a doubly negative manner. That is, participants tended to describe the consequences that would *not* occur by *not* taking the focal action. Beyth-Marom et al. (1993) hypothesized that this asymmetry reflects the difficulties individuals have in thinking about negated events. In the context of the present research, this hypothesis suggests that individuals may be more error-prone in reasoning about the prospect of no attack over differing timeframes than in reasoning about the prospect of an attack. Indeed, if people lose track of the negation, one might expect a relatively higher percentage of monotonicity violations in refocused assessments than in direct assessments of attack. I predicted that this would be the case in the between-subjects design (Experiments 1a and 1b), but not in the high-transparency within-subjects design (Experiment 3) in which it would be difficult to lose track of the negation (i.e., the "no" in no-attack) given the salience of the affirmative complement (i.e., attack).

Experiment 1a

Overview

On March 17, 2003, President Bush gave Saddam Hussein and his sons 48 hours to leave Iraq or face "the full force and might" of the American military "at a time of our choosing." Given this clear forewarning of impending war, Experiment 1a was planned to be ready to start at the first news of the war's outbreak. The first strike began on March 20, and Experiment 1a was conducted between March 21 and March 28, during the first week of the war. At the time, as many readers may recall, both the U.S. and United Kingdom administrations had claimed with confidence that Saddam Hussein had connections to Osama bin Laden and al Qaeda and, unsurprisingly, there was also much discussion at this time of possible reprisal acts of terrorism against the U.S. and its allies in the Iraq war. The key objective of Experiment 1a was to test the predicted effects of unpacking and refocusing on participants' risk assessments within a rapidly changing sociopolitical context.

Method

Participants. One hundred thirty-five participants were recruited from the University of Victoria psychology department's participant scheduling system and received partial course credit in exchange for participating in the experiment.

Design. Experiment 1a used a 4 (Question: no-attack, attack, attack-aq, attack-no-aq) × 3 (Timeframe: 2, 4, 6 months) mixed factorial design in which timeframe was manipulated within-subjects and question was manipulated between-subjects.

Procedure. The additivity of risk assessments was examined by having participants randomly assigned in roughly equal numbers to one of four conditions in which they assessed the probability of one of the following hypotheses:

> "Within the next {2, 4, or 6} months in the U.S. there will". . .
>
> (a) "be at least one major terrorist attack."
>
> (b) "be at least one major terrorist attack that is plotted by al Qaeda."
>
> (c) "be at least one major terrorist attack that is NOT plotted by al Qaeda."
>
> (d) "NOT be at least one major terrorist attack."

Each participant provided a risk assessment for periods of 2, 4, and 6 months for the relevant hypothesis which they were assigned to judge by indicating a value in the [0, 1] probability interval. Specifically, participants were instructed as follows:

> For the questions that follow, please write the decimal value that corresponds to your probability estimate. Use as many decimal places as are necessary for you to give an accurate estimate.

Following each question, participants were provided with a blank line on which to write their response. Ordering in terms of timeframe was counterbalanced such that half of each group provided their assessments in ascending "2–4–6 months" order and the other half provided their assessments in descending "6–4–2 months" order.

Results

A preliminary analysis of participants' assessments revealed that they are not normally distributed and are positively skewed for each of the three timeframes. Given that this was the case in the subsequent experiments as well, the data in all experiments were analyzed using nonparametric statistics. Moreover, to facilitate tests of the key hypotheses, in each experiment, assessments in the no-attack condition were subtracted from 1 in order to provide refocused assessments of risk of a terrorist attack. Significance levels are reported for one-tailed tests for directional predictions and two-tailed tests for nondirectional predictions in this and subsequent experiments. For most analyses, the gamma statistic, γ_1*, is reported as a nonparametric effect-size estimator, which is robust to violations of the assumptions of population normality and homogeneity of variance (Hedges & Olkin, 1985). The gamma statistic is the standard normal deviate of the sample proportion of scores in the control group that fall below the median score of the experimental group (or an arbitrary group in cases where there is no control-treatment distinction).

Effect of timeframe. A small but significant effect of timeframe on risk assessments was observed, $\chi^2(2, N = 135) = 85.05$, $p < .01$, Kendall's $W = .31$. As shown in the row labeled "Overall" in Table 1, the mean and median probabilities increased across timeframes of 2, 4, and 6 months. The mean Pearson correlation among the three risk assessments is .88.

Effect of unpacking. The unpacking hypothesis predicts that unpacked assessments (based on the sum of attack-aq and attack-no-aq assessments) would be significantly greater than packed assessments. To test this prediction, the median unpacked assessment was calculated by adding the median assessment of attack-aq to participants' assessments of attack-no-aq and, conversely, by adding the median assessment of attack-no-aq to participants' assessments of attack-aq. Table 1 shows the medians for the relevant conditions. Collapsing across the attack-aq and attack-no-aq conditions, these adjusted values thus comprise the unpacked assessments. Table 2 shows the median values of the unpacked and packed assessments, the results of the Mann–Whitney tests of the difference between these values, the effect-size estimates, and the unpacking factors (i.e., the ratio of the median unpacked assessment to the median packed assessment) as a function of timeframe. As predicted, a significant, albeit statistically small, effect of unpacking on risk assessments was observed, such that the median unpacked assessments are greater than the median packed assessments at each timeframe.

Effect of refocusing on additivity. The refocusing hypothesis predicts that the median refocused assessment would be significantly greater than the median packed assessment—namely, that

Table 1
Central Tendency of Risk Assessments as a Function of Timeframe, Question, and Experiment

	Timeframe (months)					
	2		4		6	
Question	*Mdn*	*M*	*Mdn*	*M*	*Mdn*	*M*
Experiment 1a						
1 − No-attack	.50	.58	.60	.60	.60	.62
Attack	.10	.22	.20	.29	.30	.35
Attack-aq	.10	.19	.20	.25	.30	.34
Attack-no-aq	.04	.20	.08	.23	.18	.26
Overall	.20	.30	.30	.34	.40	.39
Experiment 1b						
1 − No-attack	.90	.61	.80	.63	.90	.68
Attack	.0001	.02	.0001	.04	.003	.07
Attack-aq	.0001	.05	.0001	.10	.001	.13
Attack-no-aq	.00001	.05	.0004	.06	.01	.08
Overall	.001	.19	.02	.22	.03	.25
Experiment 2						
1 − No-attack	.12	.31	—	—	—	—
Attack	.01	.06	—	—	—	—
Experiment 3						
1 − No-attack	.05	.12	.10	.17	.13	.22
Attack	.01	.09	.04	.12	.05	.17

Note. No-attack = probability of no terrorist attack; Attack = probability of a terrorist attack; Attack-aq = probability of a terrorist attack by al Qaeda; Attack-no-aq = probability of a terrorist attack not by al Qaeda.

Table 2
Mann–Whitney Tests of the Difference Between Unpacked and Packed Assessments as a Function of Timeframe and Experiment

Timeframe (mo.)	Assessment (*Mdn*)				
	Unpacked	Packed	*U*	γ_1*	UF
		Experiment 1a			
2	.14	.10	911.0*	0.04	1.4
4	.28	.20	908.5*	0.18	1.4
6	.48	.30	784.5**	0.25	1.6
		Experiment 1b			
2	.00011	.0001	819.0**	0.72	1.1
4	.0005	.0001	857.5**	4.00	5.0
6	.011	.003	736.5**	0.96	3.7

Note. In calculating γ_1*, the packed-assessment condition was treated as the control group. UF = unpacking factor.
* $p < .05$. ** $p < .01$.

on average, 1 – *P*(no-attack) > *P*(attack). Table 3 shows the median values of the refocused and packed assessments, the results of the Mann–Whitney tests of the difference between these values, the effect-size estimates, and the sum of the median assessments attack and no-attack as a function of timeframe. As predicted, the median refocused assessments are significantly greater than the median packed assessments at each timeframe. Moreover, the effect size estimates indicate that these are large effects (following Cohen's, 1988, conventions). Table 3 also shows that the sum of the median assessments of attack and no-attack are invariably superadditive, as predicted.

Effect of refocusing on monotonicity. Overall, 19.3% of the sample provided nonmonotonic assessments across adjacent timeframes at least once. The proportion of monotonicity violations differed as a function of question, Kruskal–Wallis $\chi^2(3, N = 135) = 16.78, p < .01, \phi = .35$. As Table 4 shows, and in line with prediction, nonmonotonic risk assessments are more prevalent in the no-attack condition (i.e., for refocused assessments) than in the other conditions (i.e., for direct assessments).

Table 3
Mann–Whitney Tests of the Difference Between Unpacked and Refocused Assessments as a Function of Timeframe and Experiment

Timeframe (mo.)	Assessment (*Mdn*)				
	Refocused	Packed	*U*	γ_1*	SUM
		Experiment 1a			
2	.50	.10	251.0**	1.07	.60
4	.60	.20	265.0**	0.84	.60
6	.60	.30	294.5**	0.74	.70
		Experiment 1b			
2	.90	.0001	72.0**	4.00	.1001
4	.80	.0001	100.5**	4.00	.2001
6	.90	.003	94.0**	4.00	.103

Note. In calculating γ_1*, the packed-assessment condition was treated as the control group. SUM = the sum of the median assessments of attack and no-attack.
* $p < .05$. ** $p < .01$.

Discussion

The findings of Experiment 1a demonstrated the predicted effects of unpacking and refocusing on participants' assessments of the risk of a terrorist attack. The unpacking effect entailed subadditive risk assessments in which the sum of average assessed risk of attack by al Qaeda and non-al Qaeda sources was significantly greater than the average assessed risk of the extensionally equivalent category of attack. The refocusing effect entailed superadditive assessments in which the sum of the average summed risk of attack and no-attack fell short of unity. Moreover, as predicted and consistent with past findings (Beyth-Marom et al., 1993), monotonicity violations were most likely to occur among participants who provided refocused assessments than among those who provided direct assessments. Finally, the overall level of monotonicity violations was consistent with that shown in other research (e.g., Yechiam & Budescu, 2004).

Experiment 1b

Overview

The aim of Experiment 1b was to replicate the key findings of the previous experiment, and it was identical to Experiment 1a but for two features. First, it was conducted at a later phase of the war in Iraq. Second, a filtered response procedure was used to screen out participants who may have been entirely uncertain what to indicate as their risk assessment. Research by Fischhoff and Bruine de Bruin (1999; Bruine de Bruin, Fischbeck, Stiber, & Fischhoff, 2002) indicates that in studies of subjective probability, there is often a "blip" at .5 on the probability scale that is partly attributable to participants who express their epistemic uncertainty by responding "fifty-fifty" (see also Poulton, 1994). Indeed, an examination of the data from Experiment 1a reveals that the modal response for the 4-month and 6-month timeframes was .50 (3%, 11%, and 17% of the sample provided estimates of .50 for timeframes of 2, 4, and 6 months, respectively). Given that fifty-fifty responding would likely augment the effect of unpacking and attenuate the effect of refocusing in the present research, it is of

Table 4
Percentage of Participants Violating Monotonicity as a Function of Question and Experiment

Experiment	Question			
	No-attack	Attack	Attack-aq	Attack-no-aq
1a	42.4	11.4	5.7	18.8
1b	43.2	4.9	16.7	10.3
3	7.8	7.8	—	—

Note. No-attack = probability of no terrorist attack; Attack = probability of a terrorist attack; Attack-aq = probability of a terrorist attack by al Qaeda; Attack-no-aq = probability of a terrorist attack not by al Qaeda.

interest to examine whether these effects differ under conditions designed to eliminate fifty-fifty responding that is merely due to epistemic uncertainty.

Method

Participants. One hundred seventy-nine participants were recruited from the University of Victoria psychology department's participant scheduling system and received partial course credit.

Procedure. The experiment was conducted during the first two weeks of November 2003 during the guerrilla war phase when insurgents were claiming the lives of U.S. soldiers almost daily. Participants were randomly assigned in roughly equal numbers to one of the four risk assessment conditions used in Experiment 1a (i.e., attack, no-attack, attack-aq, or attack-no-aq). The experimental procedure was identical to that used in Experiment 1a except that participants were instructed that if they had "no idea" what to indicate in response a particular question, then they should check a box labeled "no idea." Forty-two participants (23%) checked this option for at least one of the three questions they were asked, and these participants were excluded from the experiment, leaving 137 participants.

Results

Fifty-fifty responding. Compared to Experiment 1a, the frequency of fifty-fifty responding was greatly reduced. Not a single participant indicated .50 as their response for the 2-month timeframe and only one participant did so for the 4-month timeframe. For the 6-month timeframe, only six participants (4%) indicated .50 as their response. Clearly then, the results of Experiment 1b cannot be explained in terms of fifty-fifty responding.

Variation in the central tendency of assessed risk across time. The virtual elimination of fifty-fifty responding in Experiment 1b was likely due in part to the use of the filtering procedure. However, it may also have been due to changes in the central tendency of assessed risk from the start of the war in Iraq in March, 2003, when Experiment 1a was conducted, to the November, 2003, timeframe, when Experiment 1b was conducted. To examine the latter hypothesis, I excluded responses of .50 by participants in both experiments. If the central tendency of assessed risk decreased from March to November, this should be evident even when all assessments of .50 are excluded. Table 5 shows the median risk assessments as a function of experiment (i.e., time) and timeframe and the corresponding tests of the difference between these estimates. For each timeframe, the median assessments decreased substantially—roughly one or two orders of magnitude—from March to November, 2003. Therefore, although the filtering procedure likely eliminated fifty-fifty responding due to epistemic uncertainty, the lower level of assessed risk in Experiment 1b also appears to be partly due to changes in assessed risk levels from March to November of 2003.

Table 5
Mann–Whitney Tests of the Difference Between Risk Assessments as a Function of Timeframe and Experiment Excluding Fifty-Fifty Responses

Timeframe (mo.)	Experiment (*Mdn*) 1a	1b	*U*	γ_1*
2	.10	.001	5,818.0**	0.39
4	.15	.02	5,538.5**	0.46
6	.25	.03	5,718.5**	0.42

Note. In calculating γ_1*, Experiment 1a was treated as the control group.
* $p < .05$. ** $p < .01$.

Effect of timeframe. Replicating the findings of Experiment 1a, risk assessments significantly differed as a function of timeframe, $\chi^2(2, N = 137) = 83.69$, $p < .01$, Kendall's $W = .31$. As shown in the row labeled "Overall" in Table 1, the mean and median probabilities increased across timeframes of 2, 4, and 6 months. The mean Pearson correlation among the three risk assessments is .92.

Effect of unpacking. The unpacking hypothesis was tested in the same manner as in Experiment 1a. As Table 1 shows, the unpacked assessments are, on average, significantly greater than the packed assessments at each timeframe. Moreover, these effects were large as assessed both by the gamma statistic and the unpacking factor.

Effect of refocusing on additivity violations. The refocusing hypothesis predicts that the median refocused assessment would be significantly greater than the median packed assessment. As Table 3 shows, the refocused assessments are, on average, significantly greater than the packed assessments at each timeframe. Table 3 also shows that, as in Experiment 1a, the sum of the median assessments of attack and no-attack were invariably superadditive. The degree of superadditivity in this experiment, however, was much more extreme than in Experiment 1a, and the statistical effect sizes were extremely large.

Effect of refocusing on monotonicity violations. Overall, 19.0% of the sample provided nonmonotonic assessments across adjacent timeframes at least once. This percentage is almost identical to that observed in Experiment 1a. Table 4 shows the percentage of these violations as a function of question. Monotonicity violations were nonindependent of question, $\chi^2(3, N = 137) = 20.98$, $p < .01$, $\phi = .39$. As predicted and replicating the findings of Experiment 1a, nonmonotonic risk assessments were more prevalent in the no-attack condition (i.e., for refocused assessments) than the other conditions (i.e., for direct assessments).

Discussion

Experiment 1b demonstrated that the unpacking effect and the refocusing effect were replicable even when fifty-fifty responding was eliminated and when the magnitude of the assessed probabilities were, on average across participants and assessment conditions, one to two orders of magnitude smaller than in Experiment 1a at each timeframe. Indeed, the change in the average level of assessed risk between the two experiments is itself a striking finding, which suggests that risk assessments are influenced by changing perceptions of the sociopolitical landscape. Replicating the findings of Experiment 1a, roughly one fifth of participants violated monotonicity in their assessments across timeframes, and such violations were once again most likely among participants who provided refocused rather than direct risk assessments.

Experiment 2

Overview

The preceding experiments clearly demonstrated strong, superadditive refocusing effects in between-subjects designs. The

key objective of Experiment 2 was to better understand the basis of this effect. As noted earlier, within-subject designs are better suited to differentiate coherence violations due to comprehension errors (i.e., not understanding the relevant principle) from those due to application errors (i.e., not accessing the relevant principle). Building on past research indicating that consider-the-opposite interventions tend to reduce judgment bias, I predicted that within-subject tests of the refocusing effect would replicate the effect if the relation between complementary attack and no-attack hypotheses was low in transparency, but not when this relation was relatively high in transparency. I tested this hypothesis in two steps. In Experiment 2 (low transparency), I sought to replicate the superadditive refocusing effect using a within-subject design in which the relevant complementary judgment tasks were embedded in sets of other probability assessment tasks and separated by an unrelated task (i.e., completing Cacioppo, Petty, & Kao's, 1984, Need for Cognition [NFC] scale). In Experiment 3 (high transparency), I sought to debias judgment, and thereby reduce incoherence in risk assessments, by presenting the complementary hypotheses in close succession.

Experiment 2 also had two subsidiary objectives. First, I tested the generalizability of the refocusing effect by examining whether it was replicable using a second binary complement unrelated to the risk of terrorism. Second, I examined whether the degree of coherence exhibited by participants in their risk assessments was correlated with their need for cognition (Cacioppo & Petty, 1982). Few studies have examined the relation between need for cognition and the frequency or magnitude of violations of coherence principles in judgment or choice *within* participants. Among the published studies that have examined this question, the evidence is mixed. LeBoeuf and Shafir (2003) found that high-NFC participants were less likely to violate the descriptive invariance principle in framing tasks than low-NFC participants. By contrast, Levin, Gaeth, Schreiber, and Lauriola (2002) found that NFC was not correlated with violations of invariance due to framing (also see Windschitl et al., 2003). Experiment 2 further explored this research question by examining how NFC relates to a different type of coherence violation—nonadditive probability judgments.

Method

Participants. Forty-eight participants were recruited from the University of Victoria psychology department's participant scheduling system and received partial course credit.

Procedure. Participants took part in Experiment 2 during the last two weeks of February 2004. Participants were first asked to provide seven probability assessments in decimal form using the unfiltered response procedure introduced in Experiment 1a. Having replicated the refocusing effect with the filtering procedure, the unfiltered procedure was employed in Experiment 2 because it is less susceptible to sampling bias (see Krosnick, 1999). The first and last three items were distracter items (e.g., "What do you think the probability is that Israel and the Palestinian Authority will achieve peace in the next 5 years?"). The fourth item was the target item: "What do you think the probability is that there will be at least one major terrorist attack in the U.S. within the next 2 months?"

After completing the first set of assessments, participants completed Cacioppo et al.'s (1984) 18-item NFC scale as a distracter task. Then, participants were asked to provide seven new probability assessments. Each of the assessed events was described in terms of a marked negation (e.g., "What do you think the probability is that the Palestinians will NOT have their own state within the next decade?"). Once again, the first and last three items comprised distracter items. The fourth item was the target question, which was the binary complement of the fourth item from the first set of items: "What do you think the probability is that there will NOT be at least one major terrorist attack in the U.S. within the next 2 months?" Finally, one of the distracter pairs—namely, the seventh item in the first and last set formed another binary complement. In the pre-NFC set, participants were asked to assess the probability that the Canadian dollar would increase in value relative to the U.S. dollar in 6 months (increase), whereas in the post-NFC set they were asked to assess the probability that the Canadian dollar would NOT increase in value relative to the U.S. dollar in the same timeframe (no-increase). Inclusion of this second binary complement permitted the key test of the refocusing hypothesis to be replicated using a change in content.

Results

Effect of refocusing. Replicating the between-subjects effects in the previous experiments, the refocused assessments are, on average, significantly greater than the packed assessments (see Table 1 for means and medians), Wilcoxon's $z = 4.74$, $p < .01$, $\gamma_1^* = 1.39$. The sum of the median assessments of attack and no-attack was .89, thus replicating the finding of superadditivity observed in the preceding experiments. To examine the percentage of participants that provided additive risk assessments, participants' assessments of attack and no-attack were summed. Sixty percent (95% CI = 13.86) of participants provided additive assessments using a criterion of 1 ± 0.1.

Replication with change in content. The refocusing effect was also replicated when the content of the relevant hypotheses was changed (i.e., increase vs. no-increase). In line with the previous findings, the refocused assessments ($Mdn = .50$) are, on average, significantly greater than the packed assessments ($Mdn = .10$), Wilcoxon's $z = 4.75$, $p < .01$, $\gamma_1^* = 1.53$. The sum of the median assessments of attack and no-attack was .60, once again revealing superadditivity of binary complements. Fifty percent (95% CI = 14.15) of participants provided additive assessments using the criterion of 1 ± 0.1.

Need for cognition. I correlated the NFC scale ($\alpha = .86$) with the signed and absolute deviations of the summed assessments from additivity (i.e., $1 - \Sigma$ and $|1 - \Sigma|$, respectively). Consistent with Levin et al.'s (2002) findings, none of these correlations was significant, greatest $|r| = .19$, *ns* However, given that power to detect significant correlations in this experiment is relatively low, these findings should be interpreted with caution.

Discussion

Experiment 2 replicated the superadditive refocusing effect in a within-subjects design and with a change in the content of the assessed complementary hypotheses. The statistical effects observed were large by conventional standards, and refocusing led to a five- to 12-fold increase in the assessed probability of the focal hypothesis. Because Experiment 2 used a within-subjects design, the findings also demonstrate coherence violations more directly than Experiments 1a and 1b, which relied on aggregating the assessments from different groups of participants—a potential limitation to which other research (e.g., Idson et al., 2001; Macchi et al., 1999) is also susceptible.

Experiment 2 also revealed that there are individual differences in the coherence of assessed probabilities. In spite of the strong,

superadditive refocusing effects observed at the nomothetic level of analysis, 50%-60% of participants provided roughly additive assessments. Using a 95% confidence interval, one may predict that the refocusing effect observed in the present context would be attributable to 26%-53% of the population (for the terrorism items). Individual difference factors that may account for this variability could be examined in future research. In Experiment 2, I examined one potential candidate—need for cognition. However, this measure was not significantly related to the degree of incoherence exhibited in probability judgment.

Experiment 3

Overview

As noted earlier, Experiment 3 was designed to test the hypothesis that the refocusing effect would be attenuated, if not eliminated, when participants were presented with complementary hypotheses to assess in a relatively transparent manner. Support for this hypothesis would indicate that the refocusing effect is the result of an application error rather than a comprehension error. Moreover, if that were the case, it would also indicate that structuring assessment tasks in ways that highlight the salience of counterexamples could serve as an effective debiasing strategy for probability judgment. In a related vein, I predicted that the rate of monotonicity violations in assessing no-attack would be reduced in this experiment due to the fact that it should be relatively difficult to lose track of the negation when the affirmative complement is salient.

Method

Participants. Fifty-one participants were recruited from the University of Victoria psychology department's participant scheduling system and received partial course credit.

Design and procedure. Participants took part in this experiment, as in Experiment 1b, during the first two weeks of November 2003. Participants were asked to provide six risk assessments in a 2 (Assessment: attack, no-attack) × 3 (Timeframe: 2, 4, 6 months) within-subjects design. Order of assessments (attack-first, no-attack-first) and order of timeframe (ascending, descending) were counterbalanced across participants. The six risk questions were presented on a single page that allowed participants to easily assess their relation to each other. Participants either provided assessments for the three attack questions first followed by the corresponding no-attack questions, or vice versa. The attack and no-attack questions were phrased exactly as in the previous experiments. Instructions were identical to those used in Experiment 1a. As in the preceding experiments, participants were provided with a blank line on which to write their response after each question.

Results

Effect of timeframe. Packed assessments significantly differed as a function of timeframe, $\chi^2(2, N = 51) = 64.73$, $p < .01$, Kendall's $W = .64$. Likewise, there was a significant effect of timeframe on the refocused assessments, $\chi^2(2, N = 51) = 58.97$, $p < .01$, Kendall's $W = .58$. As shown in Table 1, the mean and median packed and refocused assessments increased across timeframes of 2, 4, and 6 months. The mean Pearson correlation was .87 among the three packed assessments and .92 among the three refocused assessments.

Effect of refocusing. In order to directly compare the effect of refocusing to that observed in Experiment 2, I analyzed the comparable 2-month assessments. By contrast with Experiment 2, and in support of the application-error hypothesis, the packed and refocused assessments did not significantly differ, Wilcoxon's $z = 1.25$, *ns*, $\gamma_1^* = 0.12$ (see Table 1 for means and medians). Moreover, 82.4% (95% CI = ±10.45) of participants provided additive (i.e., 1 ± 0.1) assessments for the 2-month timeframe. This percentage is significantly greater than the percentage observed in Experiment 2, $p < .05$ by Fisher's exact test, = .24.

Violations of monotonicity. As predicted, and in contrast to the findings of Experiments 1a and 1b, the percentage of monotonicity violations in Experiment 3 was low and exactly the same for assessments of no-attack and assessments of attack (see Table 4).

Discussion

The findings of Experiment 2 supported the application-error hypothesis. If participants simply did not understand the additivity property, one might expect a degree of incoherence in judgment comparable to that observed in Experiment 2. However, that was not the case. Rather, the refocusing effect was eliminated in the present experiment, with a significantly greater proportion of the sample providing additive assessments. Moreover, providing convergent support for the application-error hypothesis, monotonicity violations in assessments of no-attack were dramatically reduced and were identical in frequency to those made in direct assessments of attack. Taken together, the findings tentatively support the idea that individuals can be distracted from considering the relevance of this consistency principle when their attention is directed toward assessing the probability of only one side of a binary partition. Finally, the findings of Experiment 3 also rule out the possibility that the refocusing effects observed in the preceding experiments were merely the result of a conversational bias in which the paired hypotheses were in fact not understood to be complementary.

General Discussion

The present research examined the coherence (i.e., the additivity and monotonicity) of risk assessments regarding the threat of a terrorist attack as a function of variations in event description (unpacking) and question wording (refocusing). The findings provided strong support for the overarching prediction that the coherence of risk assessments is susceptible to variations in both of these factors. Experiments 1a and 1b revealed that unpacking the description of a terrorist attack into sources due to al Qaeda and those other than al Qaeda produced an inflation of the subjective probability of such an attack. This inflation of subjective probability is in violation of the additivity property and indicates that, on average, participants' judgments are subadditive. These findings extend support theory's prediction of an unpacking effect (Rottenstreich & Tversky, 1997; Tversky & Koehler, 1994) to the domain of risk assessment involving the asymmetric threat of terrorism. The unpacking effect was observed using both filtered and unfiltered response modes and was evident both at the start of the war in Iraq and 8 months later during the insurgency phase of the war.

The present findings also provided strong support for the predicted superadditive refocusing effect. Experiments 1a, 1b, and 2

revealed that inferring the probability of a terrorist attack from participants' assessments of the probability of the complementary possibility of no attack led to a substantial inflation in subjective probability. This inflation of subjective probability is also in violation of the additivity property and indicates that, on average, participants' judgments are superadditive. That is, the summed probabilities of attack and no attack were less than unity. Although superadditivity has been demonstrated in past research, the degree of incoherence in this regard exceeds that observed in earlier studies (e.g., Idson et al., 2001; Macchi et al., 1999; Moore & Kim, 2003; Sloman et al., 2004; Windschitl et al., 2003). Most of the earlier studies also used between-subjects designs. Given that the average probability of *X* from one sample and the average probability of not-*X* from a different sample are not logically required to be additive, Experiment 2 offered a more conservative demonstration of the nonadditivity of binary complements than those observed in past studies.

The findings of Experiments 2 and 3 further helped to resolve the issue of whether the refocusing effect is primarily the result of a comprehension error or an application error by showing that a strong refocusing effect was observed when the transparency of the assessed complements was relatively low (Experiment 2) but not when transparency was relatively high (Experiment 3). This variation as a function of transparency is inconsistent with the comprehension-error interpretation because performance should not be affected by factors that increase the accessibility of a normative principle if that principle has not been acquired in the first place. By contrast, the findings support the application-error interpretation by showing that performance can be improved when the accessibility of the principle is increased through changes in task structure.

Finally, the present findings revealed a systematic effect of question wording on the monotonicity of risk assessments across timeframes. In Experiments 1a and 1b, violations of monotonicity were more frequent among participants who were asked to assess the risk of no attack taking place than among participants who were asked to assess the risk of an attack in any of the three variant conditions. These findings support Beyth-Marom et al.'s (1993) hypothesis that risk assessments of indirect events pose greater difficulties for individuals than those involving direct assessments. The findings of Experiment 3, however, also showed that the effect of negation on violations of monotonicity is eliminated when the complementary relation between the possibility of an attack and its complementary possibility are made salient. Taken together, the findings of Experiments 1a, 1b, and 3 supported the prediction that it is easier to lose track of a negation when the complementary affirmative statement is not directly accessible. These finding thus lend further support to the application-error interpretation by showing that factors that increase the likelihood of coherence violations can be controlled through a restructuring of tasks. If the effect of such factors on coherence were due solely to failures to comprehend the relevant normative principles, we would not expect to see such changes in effect as a result of task restructuring.

Implications for Theory on Subjective Probability

Are binary complements additive? An axiom of support theory is that probability judgments of binary complements will be additive. The empirical evidence at the time the theory was devised was consistent with this account (see Tversky & Koehler, 1994), and some recent studies continue to find evidence of additivity for binary complements (e.g., Brenner & Rottenstreich, 1999; Juslin et al., 2003). However, as noted earlier, other studies have reported superadditivity for binary complements. The apparent inconsistency between these studies may be resolved by considering the following points. First, studies vary in the procedures used to aggregate assessments for analysis. Studies that have found additivity have not only averaged probability judgments across participants, they have also averaged additivity values over a set of binary complements (e.g., Juslin et al., 2003; Wallsten et al., 1993). Given that superadditivity is most likely to be observed in cases in which complementary hypotheses each recruit little support (Idson et al., 2001; Macchi et al., 1999), aggregating across stimuli that vary in support is likely to dilute the refocusing effect. Indeed, no study reporting violations of additivity for binary complements (including this research) has indiscriminately averaged across stimuli.

Second, studies motivated by support theory tend to describe probability assessment in terms of an *evaluation frame* in which a focal hypothesis, *X*, is judged in relation to a residual hypothesis, *Y* (Tversky & Koehler, 1994). Researchers have thus elicited probability assessments by making the evaluation frame explicit, such as asking, "What is the probability that there will be a terrorist attack in the U.S. rather than no terrorist attack in the U.S.?" When the evaluation frame entails binary complements, the framing of the question will likely draw participants' attention to the complementarity relation, and thus increase the likelihood that the assessments will be additive (e.g., Brenner & Rottenstreich, 1999). Although support theory posits that people spontaneously adopt evaluation frames, the assumption remains untested and is inconsistent with conversationally pragmatic accounts (e.g., Hilton & Slugoski, 2000), which suggest that normative pressures for conversational efficiency that would militate against such an expansion.

Third, variations across studies in the transparency of the relation between hypotheses comprising a binary complement may explain apparent discrepancies in findings. It is noteworthy that studies demonstrating the additivity of binary complements in within-subjects designs have tended to use the consecutive judgment format (e.g., see Brenner & Rottenstreich, 1999). The present research clearly demonstrated that under such conditions the refocusing effect is eliminated. Thus, studies employing such designs cannot be regarded as conservative tests of the additivity of binary complements axiom.

Underlying processes. Although the aim of the present research was to demonstrate factors that may influence the coherence of risk assessments with a view to the applied implications for gauging performance in contexts where accuracy may be difficult or impossible to gauge directly, it is worth reflecting briefly on the underlying processes that might have given rise to the observed effects. The standard interpretation of the unpacking effect is that it is due primarily to the availability heuristic (Tversky & Koehler, 1994). According to this account, the more that a hypothesis is unpacked, the easier it is to bring to mind ways in which the implicit hypothesis might occur. Thus, support for the hypothesis is increased and, accordingly, subjective probability as well. Other processes, however, could also be implicated. For instance, the mere accessibility of specific ways in which an implicit hypothesis

might occur, especially if they are familiar (Koriat & Levy-Sadot, 2001), could increase subjective probability independent of using the availability heuristic, which relies more specifically on the notion that ease of recall serves as a proxy for subjective probability. In cases in which participants are fairly uncertain what probability to assign to a hypothesis, they may be likely to exhibit a response contraction bias in which they estimate a value close to .50 (Poulton, 1994). This response bias would be expected to produce subadditivity as a function of unpacking. The findings of Experiment 1b, however, militate somewhat against this interpretation because participants were given the option of responding "don't know." Still, filtering cannot fully rule out the possible effect of response contraction bias because uncertain participants may simply adjust their assessments toward .50 without indicating that value as their estimate, and this too would lead to subadditivity as a function of unpacking.

While the present research suggests that the refocusing effect is largely due to an application error (among some participants), more research would need to be conducted in order to better understand the basic underlying processes. Past research demonstrating superadditivity of binary complements suggests that this coherence violation is more likely to occur when support for both complements is low (Idson et al., 2001; Macchi et al., 1999). Given that the present research did not assess support directly or vary hypothesis in a manner that would likely change the degrees of support recruited for a given hypothesis and its complement, it cannot address this basic research question directly. Nevertheless, the findings of Experiments 2 and 3 suggest that the effect may be due to a variant of focalism in judgment (Moore & Kim, 2003; Windschitl et al., 2003) in which the degree of support for a complementary hypothesis is not considered unless it is specifically brought to the participant's attention at the time the focal hypothesis is being assessed.

Implications for Risk Assessment

The present research extends previous work on violations of coherence in probability judgment by focusing on risk assessments within an important sociopolitical domain. Previous studies examining the additivity of probability assessments have used either almanac-type questions in which the assessments elicited can be viewed as measures of participants' confidence in their answers or conditional scenarios in which participants must first assume a given future state of the world and then assess the probability of one of two possible outcomes that are described. For instance, Macchi et al., (1999) had participants answer almanac-type questions such as, "What is the probability that the freezing point of {gasoline, alcohol} is greater than that of {alcohol, gasoline}?" Idson et al. (2001) had participants provide conditional assessments such as, "Assuming that Morocco and Kenya make it to the next World Cup Finals, what is the probability that {Morocco, Kenya} wins?" By contrast, participants in the present research were asked for direct assessments of the risk of an asymmetric threat with reference to a set of well-specified and relevant timeframes. Of course, the present research focused primarily on assessments of only one type of risk. Although it is possible that the findings would not generalize to other domains, the replication of the refocusing effect in Experiment 2 using unrelated items (i.e., probability assessments regarding the direction of change in the relative value of two familiar currencies) suggests that the effect is fairly robust.

How should the practical significance of the present findings be interpreted? Quite clearly, the present findings show that the internal consistency of risk assessments can be easily perturbed. As noted earlier, coherence violations are of practical value because, while it is true that coherent judgments do not imply accurate judgments, incoherent judgments do imply that, at best, only one of the judgments elicited may be accurate. Thus, the factors influencing coherence in the present research do have an effect on forecasting accuracy even though accuracy itself could not be gauged directly. The factors identified may also be of use in risk management. As Kunreuther (2002) noted, framing can be used to make people pay more attention to small probabilities that might otherwise be dismissed. Consider a smoker who estimates his personal risk of death from smoking to be negligible. Refocusing this individual's attention on the probability of not dying from smoking might increase perceived risk if the individual is forced to confront a possible inconsistency. More generally, factors that influence the coherence of risk assessments could be used strategically to increase the probability assigned to underweighted risks and decrease the probability assigned to overweighted risks.

Extending past work on debiasing through the use of counterexamples (e.g., Hoch, 1985; Koriat et al., 1980; Wason, 1960), the present findings also suggest that the coherence of risk assessments can be increased by eliciting risk assessments in complementary ways that require a reconciliation of the assessments. For example, intelligence analysts who focus on the risk of a specific type of attack could check the stability of their assessments by reframing the task in terms of assessing the probability of that specific type of attack not occurring. To the extent that the assessments are incoherent, they might then be asked how much each of the two assessments would need to be adjusted so that they meet the requirements of extensional logic. Future research could examine such procedures in contexts where accuracy could also to be gauged in order to determine whether such procedures do in fact improve performance. For example, Mandel (2005) has shown recently that, when the probability of success is high, participants are less accurate in judging the probability of exactly one success on four independent trials than in judging the probability of exactly three failures on the four trials in spite the extensional equivalence of these descriptions. Conversely, when the probability of success is low, participants were less accurate when presented with a hypothesis about an exact number of failures rather than with one about the complementary number of successes.

Another question of practical significance is whether expertise would militate against coherence violations. There is a substantial number of studies showing that expertise improves performance in judgment tasks (for a summary, see Poulton, 1994, chap. 15). However, there is also some recent evidence that experts are highly susceptible to unpacking effects. For instance, Tetlock and Lebow (2001) showed that historians who were asked to unpack counterfactual alternatives to Western domination over an 850-year timeframe provided probability assessments of the possibility of alternatives to Western domination that were considerably greater than the assessments provided by historians who did not first unpack counterfactual alternatives. Unpackers also provided mean assessments that, when summed with another group of experts' comple-

mentary mean assessments of the inevitability of Western domination, were subadditive.

Building on the present findings, future research could examine description-related effects on risk assessments of experts, such as intelligence analysts and operational planners, who are charged with the responsibility for gathering information and assessing the risk of terrorist threats. Moreover, future research could examine whether similar coherence violations are evident in verbal as opposed to numeric risk assessments. Despite policy recommendations to communicate risks with numbers rather than words (e.g., Fischhoff, in press), military and national security doctrine on risk management often relies on verbal probability assessment procedures (e.g., Canada Department of National Defence, 2002). Future research could also examine the coherence of vulnerability assessments, which may be as important as risk assessments for managing asymmetric threats such as the possibility of terrorist attacks (Kunreuther, 2002; Slovic, 2002). For instance, would individuals assess the vulnerability of a city to attack as being equal to its invulnerability [or security] from attack subtracted from "maximum vulnerability" on the relevant scale? Answers to such questions could make useful contributions to policy development and practices aimed at managing risk and improving domestic and international security.

References

Ayton, P. (1997). How to be incoherent *and* seductive: Bookmakers' odds and support theory. *Organizational Behavior and Human Decision Processes, 72,* 99–115.

Beyth-Marom, R., Austin, L., Fischhoff, B., Palmgren, C., & Jacobs-Quadrel, M. (1993). Perceived consequences of risky behaviors: Adults and adolescents. *Developmental Psychology, 29,* 549–563.

Brenner, L. A., & Rottenstreich, Y. (1999). Focus, repacking and the judgment of grouped hypotheses. *Journal of Behavioral Decision Making, 12,* 141–148.

Bruine de Bruin, W., Fischbeck, P. S., Stiber, N. A., & Fischhoff, B. (2002). What number is "fifty-fifty"?: Redistributing excessive 50% responses in elicited probabilities. *Risk Analysis, 22,* 713–723.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42,* 116–131.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48,* 306–307.

Canada Department of National Defence. (2002). *Risk management for CF operations* (Joint doctrine manual B-GJ-005–502/FP-000). Ottawa, Ontario, Canada: Author.

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

Edwards, W. (1982). Conservatism in human information processing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 359–369). Cambridge, U. K.: Cambridge University Press.

Fischhoff, B. (in press). Risk perception and risk communication. In D. Kamien (Ed.), *The McGraw-Hill handbook of terrorism.* New York: McGraw-Hill.

Fischhoff, B., Bostrom, A., & Quadrel, M. J. (2002). Risk perception and risk communication. In R. Detels, J. McEwen, R. Reaglehole, & H. Tanaka (Eds.), *Oxford textbook of public health* (4th ed., pp. 1105–1123). Oxford, U. K.: Oxford University Press.

Fischhoff, B., & Bruine de Bruin, W. (1999). Fifty-fifty = 50? *Journal of Behavioral Decision Making, 12,* 149–167.

Fischhoff, B., & MacGregor, D. (1983). Judged lethality: How much people seem to know depends upon how they are asked. *Risk Analysis, 3,* 229–236.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1978). Fault trees: Sensitivity of estimated failure probabilities to problem representation. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 330–344.

Fox, C. R., Rogers, B., & Tversky, A. (1996). Option traders exhibit subadditive decision weights. *Journal of Risk and Uncertainty, 13,* 5–19.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond "heuristics and biases." *European Review of Social Psychology, 2,* 83–115.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment.* Cambridge, U. K.: Cambridge University Press.

Hammond, K. R. (2000). Coherence and correspondence theories in judgment and decision making. In T. Connolly, H. R. Arkes, & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (2nd ed., pp. 53–65). New York: Cambridge University Press.

Hedges, L. V., & Olkin, I. (1985). Nonparametric estimators of effect size in meta-analysis. *Psychological Bulletin, 96,* 573–580.

Hilton, D. J., & Slugoski, B. R. (2000). Judgment and decision making in social context: Discourse processes and rational inference. In T. Connolly, H. R. Arkes, & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (2nd ed., pp. 651–676). Cambridge, U. K.: Cambridge University Press.

Hoch, S. J. (1985). Counterfactual reasoning and accuracy in predicting personal events. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 719–731.

Huddy, L., Feldman, S., Capelos, T., & Provost, C. (2002). The consequences of terrorism: Disentangling the effects of personal and national threat. *Political Psychology, 23,* 485–510.

Idson, L. C., Krantz, D. H., Osherson, D., & Bonini, N. (2001). The relation between probability and evidence judgment: An extension of support theory. *Journal of Risk and Uncertainty, 22,* 227–249.

Juslin, P., Winman, A., & Olsson, H. (2003). Calibration, additivity, and source independence of probability judgments in general knowledge and sensory discrimination tasks. *Organizational Behavior and Human Decision Processes, 92,* 34–51.

Kahneman, D., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge, U. K.: Cambridge University Press.

Kahneman, D., & Tversky, A. (1982). On the study of statistical intuitions. *Cognition, 11,* 123–141.

Kahneman, D., & Tversky, A. (1984). Choices, values, and frames. *American Psychologist, 39,* 341–350.

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

Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 34–53.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 6,* 107–118.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50,* 537–567.

Kunreuther, H. (2002). Risk analysis and risk management in an uncertain world. *Risk Analysis, 22,* 655–664.

LeBoeuf, R. A., & Shafir, E. (2003). Deep thoughts and shallow frames: On the susceptibility to framing effects. *Journal of Behavioral Decision Making, 16,* 77–92.

Lerner, J. S., Gonzalez, R. M., Small, D. A., & Fischhoff, B. (2003). Effects of fear and anger on perceived risks of terrorism: A national field experiment. *Psychological Science, 14,* 144–150.

Levin, I. P., Gaeth, G. J., Schreiber, J., & Lauriola, M. (2002). A new look

at framing effects: Distribution of effect sizes, individual differences, and independence of types of effects. *Organization Behavior and Human Decision Processes, 88,* 411–429.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37,* 2098–2109.

Macchi, L., Osherson, D., & Krantz, D. H. (1999). A note on superadditive probability judgment. *Psychological Review, 106,* 210–214.

Mandel, D. R. (2005). *When equiprobable (even identical) events are judged heteroprobable: The interactive effect of framing and evidence on subjective probability.* Manuscript submitted for publication.

Moore, D. A., & Kim, T. G. (2003). Myopic social prediction and the solo comparison effect. *Journal of Personality and Social Psychology, 85,* 1121–1135.

Olson, C. L. (1976). Some apparent violations of the representativeness heuristic in human judgment. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 599–608.

Paté-Cornell, E., & Guikema, S. (2002). Probabilistic modeling of terrorist threats: A systems analysis approach to setting priorities among countermeasures. *Military Operations Research, 7,* 5–23.

Poulton, E. C. (1994). *Behavioral decision theory: A new approach.* New York: Cambridge University Press.

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

Sloman, S., Rottenstreich, Wisniewski, E., Hadjichristidis, C., & Fox, C. R. (2004). Typical versus atypical unpacking and superadditive probability judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 573–582.

Slovic, P. (1987). Perception of risk. *Science, 236,* 280–285.

Slovic, P. (2000). *The perception of risk.* London: Earthscan.

Slovic, P. (2002). Terrorism as hazard: A new species of trouble. *Risk Analysis, 22,* 425–426.

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,* 271–296.

Stone, E., Yates, F., & Parker, A. (1994). Risk communication: Absolute versus relative expressions of low-probability risks. *Organizational Behavior and Human Decision Processes, 60,* 387–408.

Tetlock, P. E., & Lebow, R. N. (2001). Poking counterfactual holes in covering laws: Cognitive styles and historical reasoning. *American Political Science Review, 95,* 829–843.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101,* 547–567.

Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: An account of deviations from Bayes's theorem and the additivity principle. *Memory & Cognition, 30,* 171–178.

Wallsten, T. S., Budescu, D. V., & Zwick, R. (1993). Comparing the calibration and coherence of numerical and verbal probability judgments. *Management Science, 39,* 176–190.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology, 12,* 129–140.

Windschitl, P. D., Kruger, J., & Simms, E. N. (2003). The influence of egocentrism and focalism on people's optimism in competitions: When what affects us equally affects me more. *Journal of Personality and Social Psychology, 85,* 398–408.

Yechiam, E., & Budescu, D. V. (2004). *When the clock slows down: The sensitivity of probability to time units.* Manuscript submitted for publication.

Received February 3, 2005
Revision received September 15, 2005
Accepted September 15, 2005 ■

DOCUMENT CONTROL DATA
(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document, Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's document, or tasking agency, are entered in section 8.)

Publishing: DRDC Toronto
Performing: DRDC Toronto
Monitoring:
Contracting:

2. SECURITY CLASSIFICATION
(Overall security classification of the document including special warning terms if applicable.)

UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification is indicated by the appropriate abbreviation (S, C, R, or U) in parenthesis at the end of the title)

Are risk assessments of a terrorist attack coherent? (U)
(U)

4. AUTHORS (First name, middle initial and last name. If military, show rank, e.g. Maj. John E. Doe.)

David R. Mandel

5. DATE OF PUBLICATION
(Month and year of publication of document.)

February 2005

6a NO. OF PAGES
(Total containing information, including Annexes, Appendices, etc.)

12

6b. NO. OF REFS
(Total cited in document.)

56

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of document, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Scientific Literature Published in the Journal of Experimental Phychology: Applied 2005, Vol.11, No.4, 277−288.

8. SPONSORING ACTIVITY (The names of the department project office or laboratory sponsoring the research and development − include address.)

Sponsoring:
Tasking:

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant under which the document was written. Please specify whether project or grant.)

9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)

10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document)

DRDC Toronto SL 2005−081

10b. OTHER DOCUMENT NO(s). (Any other numbers under which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on the dissemination of the document, other than those imposed by security classification.)

Unlimited distribution

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11), However, when further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited announcement

DOCUMENT CONTROL DATA
(Security classification of the title, body of abstract and indexing annotation must be entered when the overall document is classified)

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

(U) This article examined violations of coherence in forecasting the risk of a terrorist attack. In Experiments 1a and 1b, unpacking the risk of attack into Al−Qaeda or non−Al−Qaeda sources produced an increase in assessed risk (the unpacking effect), and deriving risk forecasts of attack by subtracting the probability of no attack from unity produced an even greater inflation of assessed risk (the refocusing effect). Experiments 2 and 3 revealed that the refocusing effect is partly due to a variant of focalism in judgment. Violations of extensional forecasting across timeframes were also observed in Experiments 1a, 1b, and 2. The findings demonstrate multiple violations of forecasting coherence in an important socio−political domain.

(U) Dans cet article, nous examinons les violations de cohérence dans la prévision du risque lié à une attaque terroriste. Dans les expériences 1a et 1b, le dégroupage du risque d'une attaque en sources al−Qaïda ou non al−Qaïda a produit une hausse du risque évalué (effet du dégroupage), et le fait de déduire les prévisions de risque d'attaque en soustrayant la probabilité d'aucune attaque de l'unité a engendré une inflation encore plus grande du risque évalué (effet de recentrage). Les expériences 2 et 3 ont révélé que l'effet de recentrage est attribuable en partie à une variation de focalisation dans le jugement. Des violations de la prévision extensionnelle pour les diverses périodes ont également été observées dans les expériences 1a, 1b et 2. Les résultats montrent de multiples violations de la cohérence prévisionnelle dans un important domaine sociopolitique.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

(U) Risk forecasting; coherence violations; additivity; terrorism