Available online at www.sciencedirect.com
ScienceDirect
Cognition 106 (2008) 130–156
COGNITION
www.elsevier.com/locate/COGNIT

Violations of coherence in subjective probability: A representational and assessment processes account ☆

David R. Mandel

Defence Research and Development Canada and Department of Psychology, University of Toronto, 1133 Sheppard Avenue West, P.O. Box 2000, Toronto, Ont., Canada M3M 3B9

Received 31 January 2006; revised 26 December 2006; accepted 2 January 2007

Abstract

Coherent judgment is a cardinal feature of rational cognition. Six experiments revealed systematic violations of coherence in probability judgment in which participants assigned different probabilities to mathematically equiprobable events. Experiments 1–5 revealed a *strict refocusing effect:* Compared to an occurrence frame, a non-occurrence frame resulted in higher estimates if base-rate evidence favored occurrence, lower estimates if evidence favored non-occurrence, and similar estimates if evidence supported indifference. Moreover, Experiments 5 and 6 revealed a *pessimistic bias* in which the less favorable of two equiprobable events was assigned greater probability. The findings support a Representational and Assessment Processes account (RAP) in which subjective probability is influenced by the perceived compatibility between representations of focal events and representations of evidence.

Keywords: Probability judgment; Coherence; Representation; Assessment; Compatibility; Framing; Pessimism

☆ This manuscript was accepted under the editorship of Jacques Mehler.
E-mail address: mandel@psych.utoronto.ca

0010-0277/$ - see front matter
doi:10.1016/j.cognition.2007.01.001

1. Introduction

Internal consistency or coherence of judgment is a cardinal feature of human rationality. A key research area addressing this important issue is the study of how well humans conform to the logic of probability calculus in judging probability (e.g., Gilovich, Griffin, & Kahneman, 2002; Kahneman, Slovic, & Tversky, 1982). A fundamental principle in this regard called *extensionality* requires that events sharing the same set-theoretic extension be assigned the same degree of probability. Many other coherence principles (e.g., monotonicity and additivity) are special cases of extensionality. Extensionality also implies descriptive invariance – namely, that the probability assigned to an event should not depend on the manner in which the event is described or "framed" (Tversky & Kahneman, 1981). That is, given that an event must be coextensive with itself across descriptions, assigned probabilities must also reflect this fact if they are to be regarded as coherent. Stated differently, if two or more descriptions of an event are truly synonymous, they should be assigned the same degree of probability.

There is ample evidence, however, that the probabilities people assign to events are influenced by the manner in which those events are described. For instance, several studies have shown that the probability assigned to a "packed" description of a disjunction, $P(X)$, tends to be less than that assigned to its "implicitly unpacked" equivalent, $P(X_1 \vee X_2 \cdots \vee X_n)$, which in turn tends to be less than that assigned to its "explicitly unpacked" equivalent, $P(X_1) + P(X_2) \cdots + P(X_n)$ (e.g., Ayton, 1997; Fischhoff, Slovic, & Lichtenstein, 1978; Koehler, Brenner, & Tversky, 1997; Mandel, 2005; Rottenstreich & Tversky, 1997; Teigen, 1983; Wright & Whalley, 1983). Such findings have led many theorists to conclude that "probability judgments are attached not to events but to descriptions of events" (Tversky & Koehler, 1994, p. 548).

The present article develops a Representational and Assessment Processes account (RAP for short) that builds on the insight that probability is assigned to event descriptions by proposing that such descriptions as well as evidence must be mentally represented before probability is assigned. These representations form the primary inputs into an assessment process that constitutes the proximal determinant of subjective probability. Although the assessment process in probability judgment may be influenced by a variety of factors (e.g., the accessibility of information and knowledge from memory, an individual's degree of personal involvement in a judgment task, or the level of cognitive load under which an individual is operating), RAP proposes that the process itself is primarily attuned toward assessing the perceived compatibility between representations of the focal event whose probability is being judged and representations of the available evidence.

In line with Anderson (1978), RAP does not restrict the nature of representation to a single type. Representations of events and evidence may have verbal, pictorial, and/or abstract propositional characteristics. Propositional representation is likely to be a necessary condition for probability judgment because coherence criteria in fact require truth-functionality, which events or their mere descriptions do not possess. For example, the additivity principle is often stated as requiring that the probability of X or its negation equals unity. However, as Moldoveanu and Langer (2002) point out, additivity is not logically required unless the descriptors "X" and "its negation" are propositionalized as "X is true or X is false." Because it would be conversationally infelicitous to spell out the truth values of events in everyday conversation (Hilton & Slugoski, 2000), references to events will tend to be interpreted as shorthand expressions of the propositions themselves. Probability judgment is also influenced by propositional representation because the number of plausible propositions that map onto a description can be large in many instances. For example, consider the question, "What is the probability of a terrorist attack?" An individual's propositional representation of the description "terrorist attack" must fill a great deal of missing yet relevant information, such as the event location, the timeframe for the event, and various attributes of the event class that collectively define which events are to be included or excluded as examples of terrorism. Propositional representations that narrow the relevant event class (i.e., more stringent exclusion criteria and less stringent inclusion criteria) are likely to attenuate subjective probability.

The role of mental representation in probability judgment, however, goes beyond propositionalization. Even when alternative descriptions map on to the same propositional representation, there may be other differences in the representations evoked that can undermine the coherence of judgment. For example, people are more responsive to risks that are described as occurring 10 times per 1,000 than they are to risks described as occurring 1 time per 100 (Slovic, Monahan, & MacGregor, 2000). They prefer gambles that offer a 10 in 100 chance of winning to those involving a 1 in 10 chance (Pacini & Epstein, 1999). And, they are more suspicious of an individual when he randomly selects a desired object (viz., a chocolate chip cookie) from a jar containing 1 such object out of 10 than when he selects it from a jar containing 10 such objects out of 100 (Miller, Turnbull, & McFarland, 1989). These effects are unlikely due to differences in propositional representation. In fact, Miller et al.'s participants tended to view the draws in the two conditions as equally probable. Rather, it appears that other representational features result in an over-weighting of the magnitude of the numerator provided in descriptions having a k-out-of-n form, perhaps because the events represented by the numerator are figural against the background events represented by the denominator (Pacini & Epstein, 1999).

The present article examines a type of statistical reasoning problem that has characteristics similar to the problem just discussed – namely, the alternative descriptions that are provided to participants are equiprobable, if not identical, from a propositional standpoint, yet they differentially influence probability judgment. In each of the six experiments reported, participants were asked to judge the probability of a compound event that consists of some number of occurrences (k) and a complementary number of non-occurrences ($n - k$) of an event, X, out of a particular trial length (n). In most of the experiments (viz., Experiments 1 -5), the framing of the compound event whose probability participants were asked to judge was manipulated such that it was either described as occurring exactly k times out of n or as not occurring exactly $n - k$ times out of n. Given that these complementary descriptions are biconditional, framing should not influence probability judgment if it is coherent. The findings of the present research, however, reveal that equiprobable events – and, indeed, even the same events framed differently – are often assigned quite different degrees of probability. These violations of the extensionality principle are systematic and shed light on the cognitive processes that underlie probability judgment. The next two subsections develop the theoretical background that guided this research.

1.1. The strict refocusing effect

In the present article, it is proposed that incoherence may result from varying the features of an event that are made explicit and the complementary features that remain implicit in a description. This type of framing manipulation is called *strict refocusing* to differentiate it from other types that do not rely on complementary relations between features of focal events (e.g., unpacking-repacking). To use a visual analogy, the explicit part of a description is like the part of an iceberg that is visible above the surface of the water, whereas the implicit part is like the part that is submerged. For example, describing an event as "one head in four coin tosses" versus "three tails in four coin tosses" illustrates a strict refocusing manipulation. The study of strict refocusing on probability judgment thus bears resemblance to the study of gain-loss framing effects on choice (Mandel, 2001; Tversky & Kahneman, 1981), although the psychological bases for these two types of framing effects are assumed to differ.

The term *strict* is used here to differentiate the proposed effect from a related one that Mandel (2005) called the refocusing effect, but which may be better described as a *procedural* refocusing effect. Mandel (2005) showed that the probability assigned to the occurrence of a terrorist attack in a specified timeframe was, on average, significantly lower than the probability assigned to its "no attack" complement subtracted from unity, implying that the total probability assigned to binary complements was superadditive (see also Idson, Krantz, Osherson, & Bonini, 2001; Macchi, Osherson, & Krantz, 1999; Windschitl, Kruger, & Simms, 2003). Procedural refocusing demonstrates that drawing attention to the probability of *different* but complementary events can elicit incoherent risk estimates. However, because the two descriptions do not refer to the same event, the procedural refocusing effect is not, strictly speaking, a framing effect. In the present research, the refocusing manipulation is strict in the sense that the *same* event is framed in different ways that focus on its complementary features.

According to RAP, strict refocusing will influence the coherence of subjective probability by altering the degree to which event descriptions are perceived as compatible with (or representative of) their evidentiary bases. The terms *compatibility* and *representativeness* are used interchangeably in this article to refer to the perceived match between representations of a focal class of events and representations of evidence (see also Kahneman & Tversky, 1972; Rottenstreich, Brenner, & Sood, 1999; Sloman, Rottenstreich, Wisniewski, Hadjichristidis, & Fox, 2004; Smith, Shafir, & Osherson, 1993). To illustrate the predictions of RAP, consider the following example: You are a contestant on a game show in which you play a lottery game with binary outcomes of either a success (X) or no success ($\neg X$, where $\neg$ means *not*) on each of four independent trials. On each trial, the probability of success,

p, is 0.9. The mathematical probabilities of 0, 1, 2, 3, and 4 successes correspond to that of 4, 3, 2, 1, and 0 non-occurrences of success, respectively, and are calculable using the binomial probability equation:

$$\Phi(X = k) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (1)$$

In Eq. (1), n is the total set size (in the present example, $n = 4$ trials), k is the frequency of the focal event, X (e.g., in the description, "exactly one success in four trials," $k = 1$), and p is the probability or base rate of X on any independent trial. The expression $\Phi(X = k)$ refers to the mathematical probability of X occurring k times out of n, and it is distinguished from $P(X = k)$, which is used later in the article to refer to the corresponding subjective probability.

The *focalism principle* in RAP proposes that people tend to represent only those elements of a proposition that are explicit in its description. It is supported by evidence of focalism in several task domains, include assessments of counterfactual mutability (see Kahneman & Tversky, 1982), forecasts of hedonic satisfaction (Schkade & Kahneman, 1998), and deductive reasoning (Legrenzi, Girotto, & Johnson-Laird, 1993). In particular, it is supported by recent studies showing that evidence in favor of a complementary hypothesis is not considered unless it is brought to attention at the time the focal hypothesis is being assessed (e.g., Moore & Kim, 2003; Windschitl et al., 2003). In the present context, the focalism principle predicts that by focusing attention on the occurrence of success (i.e., $X = 1$), people tend not to consider the biconditional implication that three successes must also not occur (i.e., $\neg X = 3$); and, conversely, by focusing attention on the non-occurrence of three successes, people tend not to consider the implied occurrence of one success.

The focalism principle bears some resemblance to the *truth principle* in Johnson-Laird, Legrenzi, Girotto, Legrenzi, and Caverni (1999) mental models theory of naïve probability (henceforth, MMT), which currently is the only other account that offers specific predictions regarding mental representation processes in reasoning about probabilities. The truth principle has two propositions: First, people represent, by default, only those possibilities (i.e., models) that conform to a focal event or set of events (e.g., $X = 1$) – namely, the true possibilities (see also, Johnson-Laird & Savary, 1999; Johnson-Laird, Legrenzi, Girotto, & Legrenzi, 2000). Second, people represent only what is "true" in each true possibility. The latter proposition refers to the idea that people represent only what is explicit in a description, and it is in this respect that the truth and focalism principles overlap and reflect an emphasis on the parsimony of representation (Johnson-Laird, 2006). The left side of Fig. 1 presents a representation for $X = 1$ (top panel) and $\neg X = 3$ (bottom panel) that is implied by MMT and that is consistent with RAP. However, because RAP, unlike MMT (see, e.g., assumption 2 in Johnson-Laird, 2006), does not require each true possibility to be represented as a distinct model, it also permits other representations of the focal event as long as they conform to the focalism principle. For instance, the four trials may be represented as a pie made up of four slices. In the $X = 1$ frame, there is one "Success" slice and the remainder of the pie is missing, whereas in the $\neg X = 3$

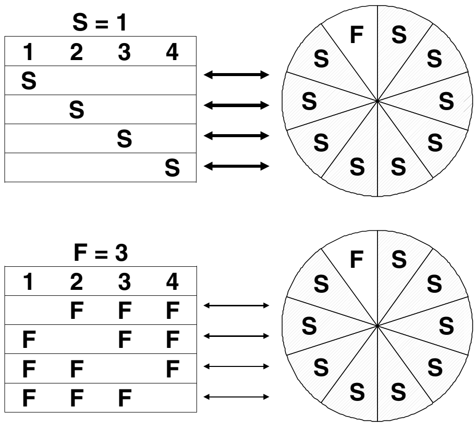

Fig. 1. Possible representation of focal compound events and evidence in the high base-rate condition of the game-show problem.

frame, there are three ‘‘Failure’’ slices and one missing slice. More generally, RAP’s elimination of the constraint that a disjunction of *n* propositions be represented by *n* models (see, e.g., Johnson-Laird, Byrne, & Schaeken, 1992, p. 424) allows for representations of disjunctions that are even more parsimonious than those permitted by MMT.

RAP differs strikingly from MMT in other key respects. According to MMT, probability assessments of $X = 1$ and $\neg X = 3$ should follow the *numerical principle*, which states that models representing the true possibilities for a focal event (say, $X = 1$) can be tagged by their numeric probabilities if such probabilities are calculable and if the possibilities for the event-generating model (e.g., for the game) are not equiprobable. MMT posits that people will then sum the probabilities of the true possibilities that conform to the focal event. Although MMT does not specify how people derive the numerical tags for particular models from more basic statistical evidence such as base rates, it is clear that in order to correctly tag the four relevant possibilities that conform to the focal event class, the probabilities of the conjoint events must be multiplied. This requires representing not only the events that are explicit in a description, but also those that are implicit. Thus, in order to accurately apply the numerical principle, one would apparently have to violate the truth principle, leading to an inconsistency in the theory. It is also unclear how the mental logic of multiplying through the probabilities of the conjoint events can be squared with MMT, which has traditionally viewed mental logic accounts of reasoning (e.g., Braine & O’Brien, 1991) as a theoretical competitor. MMT, therefore, does not predict a refocusing effect because numerical tagging would require fleshing out the conjoint events in each model of the focal event class. Of course, it is possible that people incorrectly tag the models they represent in the focal set, but the theory is silent on when such errors are probable and on how inaccuracies in numerical tagging would undermine its claim that a significant class of probability problems are reasoned through in a manner conforming to the logic of extensional reasoning.

RAP, in contrast, bypasses the need to postulate numerical tagging by proposing that the representation of evidence will be simplified in ways that would make $X = 1$ seem more probable than $\neg X = 3$. According to the *redundancy-reduction principle*, and consistent with other research indicating that reasoners favor a single representation when possible (Richardson & Ormerod, 1997; Sloutsky & Goldvarg, 2002), redundant representations will be simplified into a single representation. Given that the base rate of success is the same on each trial in the game-show problem, RAP posits that people would generate a single distributional representation of the event-generating model by, say, imaging a single wheel with 10 strips on it, nine of which are labeled S (for success occurring) and one of which is labeled F (for failure; viz., the nonoccurrence of success). As depicted in Fig. 1, the account proposes that individuals will then heuristically assess the compatibility of the possibilities corresponding to the representation of the focal outcome to that event-generating model.

According to the *compatibility principle*, the probability assigned to a focal event (e.g., $X = 1$) will be a function of the perceived compatibility of the representations of the possibilities denoted by that event and the representation of the event-generating model. Taken together, the focusing, redundancy-reduction, and compatibility principles in RAP lead to the prediction that strict refocusing will interact with the nature of evidence to systematically influence probability judgment. Specifically, a frame-by-evidence crossover interaction effect on probability judgment is predicted such that (a) an occurrence frame would be judged more probable than a non-occurrence frame if the evidence favors X's occurrence (e.g., p is high [specifically, 0.9 in the present research]) and (b) an occurrence frame would be judged less probable than a non-occurrence frame when the evidence favors X's non-occurrence (e.g., p is low [specifically, 0.9 in the present research]). Experiments 1–6 tested these and related predictions of RAP.

1.2. The pessimistic bias

Evidential representations may also be influenced by expectancies or beliefs about the distribution of various event types in daily life, which go beyond the available information that is specific to a given case. For instance, Teigen and Keren (2003, Experiment 4) found that participants believed that a lottery player with a history of wins would be less surprised by a win in a current lottery than another player who had never won but who in fact had a higher probability of winning in the same lottery. Teigen and Keren (2002, 2003, Experiment 1) also found that people are more surprised by uncontrollable successes than by uncontrollable failures even when the two types of event are described as equiprobable. This suggests that expectancies for uncontrollable negative events are stronger than for uncontrollable positive events that are nevertheless equiprobable.

The present research examined whether probability judgments are similarly susceptible to the favorability of focal events under conditions in which they are mathematically equiprobable and in which that probability can be deduced from the information supplied in the problem. RAP posits that such influences are another manifestation of the compatibility principle. That is, if the case-specific evidence indicates that two events are equiprobable, but one type of event is believed to be more probable than the other in general, then the probability assigned to the event that is more compatible with this belief will tend to be greater than that assigned to the event with which it is less compatible. Specifically, a *pessimistic bias* was hypothesized in which the less favorable of two mathematically equiprobable events is assigned greater probability. In Experiment 5, it was predicted that, when the probability of success on a given trial in the game-show problem equals the probability of no success (i.e., $p = 0.5$), participants would assign greater probability to one success than to three successes out of four trials, whereas they would assign greater probability to three failures than to one failure out of four trials, despite the equiprobability of these compound events. In Experiment 6, it was predicted that participants would assign greater probability to the less favorable of two equiprobable monetary outcomes.

2. Experiments

2.1. Experiment 1

Experiment 1 tested the hypothesis that subjective probability would be influenced by the interaction of frame (i.e., strict refocusing) and evidence (i.e., the base-rate information provided) in a manner consistent with the predictions of RAP. Specifically, it was predicted that, when the base rate was high, participants would assign greater probability to the kX (i.e., occurrence) frame than to the $(n - k)\neg X$ (i.e., non-occurrence) frame, whereas the direction of this inequality would be reversed when the base rate was low.

2.1.1. Method

2.1.1.1. Participants. The participants were 80 students from the University of Victoria who were recruited by advertisement through the psychology department subject pool and who received bonus credit in an introductory course.

2.1.1.2. Design and procedure. Participants were randomly assigned to one of four experimental conditions in a 2 (Base rate: 0.1, 0.9) × 2 (Frame: *occurrence*, *non-occurrence*) factorial design. In the description that follows, wording in square brackets and braces pertains to the 0.1 and 0.9 conditions, respectively. Participants were first told:

> Imagine that there are four events (let's just call them A, B, C, and D). Each event has exactly a [10%] {90%} chance of occurring.

Then participants were presented with the probability judgment task. In the *occurrence* condition, they were asked:

Table 1
Mean probability judgments (0–100) in Experiment 1 as a function of base rate and frame

Frame	Base rate		M
	0.1	0.9	
Occurrence	28.45	52.98	40.72
Non-occurrence	50.20	36.95	43.58
M	39.32	44.97	42.15

> What is the probability that exactly [three] {one} of those events will occur (but it could be any [three] {one} of those events)?

In the *non-occurrence* condition, participants were asked:

> What is the probability that exactly [one] {three} of those events will not occur (but it could be any [one] {three} of those events)?

Participants then indicated a value between 0 (*completely improbable*) and 100 (*completely certain*) that best reflected their judgment.[1]

2.1.2. Results

A two-way (Base rate $\times$ Frame) ANOVA indicated that only the predicted interaction effect was significant, $F(1,76) = 7.03$, $MSE = 1015.92$, $P = 0.010$. As shown in Table 1, the mean probability assigned to the occurrence frame exceeded that assigned to the non-occurrence frame when the base rate was high, but the converse was true when the base rate was low.[2] Thus, in each evidentiary condition, a strict refocusing effect was observed such that the frame predicted to be most compatible with the base-rate evidence did in fact elicit higher probability estimates than the less compatible frame.

2.2. Experiment 2

Given the abstract nature of the problem presented to participants in Experiment 1, it is important to demonstrate that the predicted interaction effect is replicable with the addition of a content-based cover story. This was the aim of Experiment 2.

2.2.1. Method

2.2.1.1. Participants. Participants were 78 students from University of Victoria recruited in the same manner as in Experiment 1.

[1] Supporting the appropriateness of these anchors, Teigen (2001) has shown that, for most participants, *improbable* signifies a low probability and *certain* signifies a high probability.

[2] Note that $\Phi(X = 1|p = 0.9) = \Phi(\neg X = 3|p = 0.9) = \Phi(X = 3|p = 0.1) = \Phi(\neg X = 1|p = 0.1) = 0.0036$.

Table 2
Mean probability judgments (0–100) in Experiment 2 as a function of base rate and frame

Frame	Base rate		*M*
	0.1	0.9	
Occurrence	18.20	60.95	39.03
Non-occurrence	60.21	32.40	48.87
M	41.59	46.31	43.95

2.2.1.2. Design and procedure. The design was identical to Experiment 1 and the only variation in procedure concerned the change in content. For example, in the 0.9 condition, participants read and answered the following:

> Imagine that there are 4 cities called Sambria, Zuwani, Mandigo, and Brallo. Each of these cities is in a different part of the world. However, in each city there is exactly a 90% chance (namely, the probability is .90) that it will be sunny on any given day during the summer.
> What is the probability that it [is sunny in exactly one] {is not sunny in exactly three} of those cities today (it could be any [one] {three} of the four cities)?
> Please indicate a value from 0 (completely improbable) to 100 (completely certain) that best reflects your assessment. My probability estimate is: ________.

2.2.2. Results

A two-way (Base rate × Frame) ANOVA indicated that only the predicted interaction effect was significant, $F(1,74) = 37.28$, $MSE = 766.04$, $P < 0.001$. As shown in Table 2, the mean probability assigned to the occurrence frame once again exceeded that of non-occurrence frame when the base rate was high, but the converse was true when the base rate was low. These findings replicate the evidence-moderated strict refocusing effects observed in Experiment 1.

2.3. Experiment 3

Experiment 3 had three objectives. The first was to replicate the predicted effect using a different content-based scenario in which the independence of outcomes across trials was even more transparent. That is, it could be argued that weather outcomes in different cities are not stochastically independent even if the cities are in different parts of the world. The second objective was to eliminate the possibility that participants were simply neglecting the term *exactly* in the probability question. Thus, a longer expression was used to ensure that participants did not fail to notice the qualifier. For instance, instead of stating "exactly three times," the question would state "no more or less than three times." The third objective was to demonstrate that the same interactive effect of base rate and frame could be observed regardless of whether participants were asked to think about the probability of *success* or to think about the probability of *failure* either occurring or not occurring.

Thus, the two factors manipulated in the previous experiments were crossed with a reflection manipulation of gains versus losses (Kahneman & Tversky, 1979). It was predicted that, consistent with the previous experiments, only the base rate by frame interaction indicative of a strict refocusing effect would be significant in the three-way model.

2.3.1. Method

2.3.1.1. Participants. Participants were 78 students from University of Victoria recruited in the same manner as in Experiment 1.

2.3.1.2. Design and procedure. Participants were randomly assigned to one of eight conditions in a 2 (Base rate: 0.1, 0.9) × 2 (Frame: *occurrence*, *non-occurrence*) × 2 (Reflection: *success*, *failure*) factorial design. In the *success* condition, participants were presented with one of the two base-rate variations of the following scenario:

> You are a contestant on a game show. You are asked to spin a wheel four times. Each time you spin the wheel there is precisely a [10%] {90%} chance that you will land on a winning strip (namely, the probability of winning on a particular spin is [.10] {.90}). If you never land on a winning strip, you win nothing. If you land on a winning strip once, you will win $10. If you land on a winning strip twice you will win $100. If you land on a winning strip three times you win will $500. And if you land on a winning strip four times you will win $1,000.

In the *failure* condition, all references to winning were changed to losing, and the appropriate changes were made to the payoff information at the end of the scenario so that the payoffs were the same as in the *success* condition. The exact wording of the probability task as a function of condition is shown in Table 3. Participants indicated a response between 0-100 using the same procedure as in the Experiments 1 and 2.

2.3.2. Results

A three-way (Base rate × Frame × Reflection) ANOVA indicated that only the predicted base rate × frame interaction effect was significant, $F(1, 67) = 20.93$, $MSE = 786.11$, $P < 0.001$ (all other Fs < 1). As shown in Table 4, the mean probability assigned to the occurrence frame once again exceeded that of non-occurrence

Table 3
Task wording in Experiment 3 as a function of base rate (p), frame, and reflection

p	Frame	Question wording "What is the probability that, no more or less than:"
0.1	Occurrence	three times in the game, you will land on a (winning, losing) strip?
0.1	Non-occurrence	once in the game, you will NOT land on a (winning, losing) strip?
0.9	Occurrence	once in the game, you will land on a (winning, losing) strip?
0.9	Non-occurrence	three times in the game, you will NOT land on a (winning, losing) strip?

Table 4
Mean probability judgments (0–100) in Experiment 3 as a function of base rate, frame, and reflection

Frame	Base rate		*M*
	0.1	0.9	
Success			
Occurrence	14.44	47.35	30.90
Non-occurrence	53.43	21.50	36.63
M	33.94	33.75	33.84
Failure			
Occurrence	22.27	48.44	34.05
Non-occurrence	51.11	23.33	37.22
M	35.25	35.89	35.55
Overall			
Occurrence	18.75	47.90	32.56
Non-occurrence	52.27	22.37	36.92
M	34.63	34.79	34.71

frame when the base rate was high, but the converse was true when the base rate was low. This pattern of findings, moreover, did not depend on whether participants were assessing success or failure. The present findings demonstrate that the interaction of frame and base-rate evidence predicted by RAP is robust. The findings replicated those of the previous experiments despite the change in how values were qualified in the probability task (i.e. "exactly" vs. "no more or less than") and the change in cover story.

2.4. Experiment 4

Having demonstrated the evidence-moderated strict refocusing effects predicted by RAP in Experiments 1–3, the objective of Experiment 4 was to examine the proposed mediating role of the perceived compatibility of descriptions to their evidentiary bases. Participants were asked to judge the representativeness of a focal description (whose probability they were also asked to judge) to the base-rate evidence that they received. It was predicted that the base rate by frame interaction effect on probability judgment would be mediated by perceived representativeness, and that a comparable two-way interaction effect on representativeness judgments would be observed.

2.4.1. Method

2.4.1.1. Participants. Participants were 228 students from University of Victoria recruited in the same manner as in Experiment 1. Data from one participant was deleted because the participant provided a probability estimate outside the requested interval.

2.4.1.2. Design and procedure. Participants were randomly assigned to one of four experimental conditions in a 2 (Base rate: 0.1, 0.9) × 2 (Frame: *occurrence*, *non-occur-*

rence) factorial design. Participants were first presented with the success version of the game-show scenario used in Experiment 3. A sub-sample of 167 participants was then presented with probability and representativeness judgment tasks, the order of which was counterbalanced across participants. The remaining 60 participants completed only the representativeness judgment task to permit a test of whether the probability judgment task had a contaminating effect on assessments of representativeness.

For the probability judgment task, participants were presented with one of the four probability questions used in the *success* condition of Experiment 3 (see Table 3), except that the expression *no more or less than* was replaced with the less cumbersome term *exactly*. Unlike the previous experiments, participants were asked to indicate their response by writing a decimal value between 0 and 1. Specifically, they were instructed, as in Mandel (2005), as follows:

> For the question that follows, 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. For example, if you think there is a 25% chance, then write .25. If you think the probability is exactly two out of every thousand, then write .002. Or, if you think the probability is exactly one in a million, then write .000001.

For the representativeness judgment task, participants first read the following instructions:

> Given the description of the game show just provided, some game show results might appear more representative or a better match to the chances of winning than others. Rate the degree to which the following result seems representative of (namely, a good match to) the chances of winning for this game. Provide your response on the scale ranging from 0 (*not at all representative*) to 10 (*extremely representative*).

Then, participants were presented with a brief statement describing the target result. This result corresponded to the base rate and frame condition to which they were assigned. For example, in the *0.9-occurrence* condition, participants assessed the result, "You play the game and you land on a winning strip <u>exactly once</u>" (underlining in original).

2.4.2. Results

Representativeness judgments did not differ as a function of judgment order or whether participants completed the probability judgment task. Therefore, the data are collapsed across these control factors in subsequent analyses.

Consistent with the findings of Experiments 1–3, a two-way ANOVA on probability judgments revealed that only the predicted base-rate × frame interaction effect was significant, $F(1, 163) = 42.64$, $MSE = 0.10$, $P < 0.001$. As shown in Table 5, the mean probability assigned to the occurrence frame was two and a half times greater

Table 5
Mean probability judgments (0–1) in Experiment 4 as a function of base rate and frame

Frame	Base rate		M
	0.1	0.9	
Occurrence	0.13	0.48	0.32
Non-occurrence	0.52	0.22	0.37
M	0.35	0.35	0.35

than that assigned to the non-occurrence frame when the base rate was high, whereas the mean probability assigned to the non-occurrence frame was four times greater than that assigned to the occurrence frame when the base rate was low. The predicted strict refocusing effects were thus replicated despite the use of a response mode that permitted precise estimates of probability.

Consistent with the preceding analysis, a two-way ANOVA on representativeness judgments revealed that only the predicted base-rate × frame interaction effect was significant, $F(1, 223) = 19.42$, $MSE = 9.11$, $P < 0.001$. As shown in Table 6, the predicted crossover interaction effect observed with probability judgments was also observed with representativeness judgments. Participants who were told that the base rate of success on each trial was low tended to judge one non-occurrence in four trials as being more representative of the evidence than three occurrences in four trials. Conversely, participants who were told that the base rate of success on each trial was high tended to judge one occurrence in four trials as being more representative of the evidence than three non-occurrences in four trials.

To directly examine the mediating role of perceived representativeness, the mediation-test procedure proposed by Kenny, Kashy, and Bolger (1998) was conducted. Fig. 2 shows the standardized regression weights and their significance levels in parentheses. As can be seen, the conditions required for mediation were met. First, the predictor (the base rate × frame interaction) significantly predicted the mediator (representativeness judgment). Second, controlling for the predictor, the mediator significantly predicted the criterion (probability judgment). Third, controlling for the mediator, the effect of the predictor on the criterion was significantly attenuated, Sobel $t = 2.39$, $P < 0.02$. These findings directly support the hypothesis that the strict refocusing effects predicted by RAP and observed in this and the preceding experiments are mediated by the perceived compatibility of the focal event and evidence.

Table 6
Mean representativeness judgments (0–10) in Experiment 4 as a function of base rate and frame

Frame	Base rate		M
	0.1	0.9	
Occurrence	4.08	5.39	4.77
Non-occurrence	5.22	3.00	4.12
M	4.71	4.16	4.43

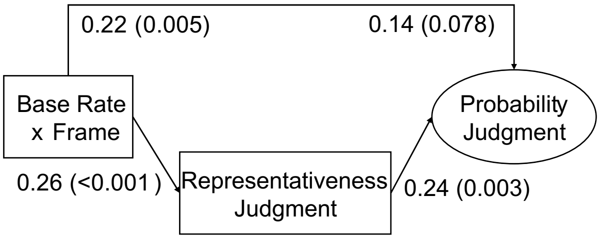

Fig. 2. Mediator model of probability judgment in Experiment 4.

2.5. Experiment 5

If strict refocusing effects are moderated by the nature of evidence (as the findings of Experiments 1–4 indicate) and if this interactive effect is further mediated by compatibility assessments (as the findings of Experiment 4 indicated), then the observed base-rate by frame interaction effect on probability judgments should be eliminated under conditions in which the base-rate evidence provided to participants was about equally compatible with the occurrence or non-occurrence of X. To test this hypothesis, an intermediate base-rate condition ($p = 0.5$) was added to the design employed in the previous experiments. Participants in this condition were randomly assigned to one of four event-frequency conditions in which they were asked to assess the probability of either $X = 1$, $X = 3$, $\neg X = 1$, or $\neg X = 3$. Once again, a significant base-rate by frame interaction effect was predicted. However, the prediction was more refined than in the previous experiments because the simple effect of frame was predicted to be nonsignificant in the intermediate 0.5 condition.

Inclusion of the intermediate condition also permitted the predictions of MMT to be pitted against the predicted pessimistic bias. In the 0.5 condition, all of the 16 (i.e., 2^4) possibilities for the event-generating model in the game-show problem are equiprobable (i.e., $0.5^4 = .0625$). Under this condition, MMT proposes that the probability of the focal event will be derived by dividing the number of models in that class by the total number of models. According to this *proportionality principle*,

$$\begin{aligned}\overline{P}(X = 1|p = 0.5) &= \overline{P}(X = 3|p = 0.5) = \overline{P}(\neg X = 1|p = 0.5)\\ &= \overline{P}(\neg X = 3|p = 0.5) = 4/16 = 0.25. \end{aligned} \tag{2}$$

That is, MMT predicts that the mean probabilities in all four event-frequency conditions will not significantly differ from one another and, moreover, that they ought not to differ from the value of 0.25. In contrast, the predicted pessimistic bias implies a significant frame by event-frequency interaction effect such that

$$\overline{P}(X = 1|p = 0.5) < \overline{P}(X = 3|p = 0.5) \tag{3}$$

$$\overline{P}(\neg X = 1|p = 0.5) > \overline{P}(\neg X = 3|p = 0.5). \tag{4}$$

That is, probability estimates of successes are expected to decrease from $k = 1$ to $k = 3$, whereas probability estimates of non-occurrences of success are expected to increase from $k = 1$ to $k = 3$.

2.5.1. Method

2.5.1.1. Participants. Participants were 149 students from University of Victoria recruited in the same manner as in Experiment 1.

2.5.1.2. Design and procedure. Participants were randomly assigned to one of six experimental conditions in a 3 (Base rate: 0.1, 0.5, 0.9) × 2 (Frame: *occurrence*, *non-occurrence*) factorial design. Participants were presented with the *success* version of the game-show scenario used in Experiment 3 and then completed the probability judgment task following the procedure described in Experiment 4. The only difference was that participants in the 0.5 condition were asked, ''What is the probability that exactly (once, three times) in the game, you [will] {will NOT} land on a winning strip?'' Moreover, in the 0.5 condition, half of the participants in the *occurrence* and *non-occurrence* conditions were asked to judge the probability of one occurrence or one non-occurrence and the other half were asked to judge the probability of three occurrences or three non-occurrences, respectively. This nested factor is referred to as *event frequency*.

2.5.2. Results

A two-way ANOVA revealed that only the predicted base-rate × frame interaction effect on probability judgments was significant, $F(2,142) = 36.76$, $MSE = 0.09$, $P < 0.001$. As shown in Table 7, the findings support the hypothesis that perceived compatibility mediates this interaction. First, the findings replicate those of the previous experiments: when the base rate was low, greater probability was assigned to the non-occurrence frame than to the occurrence frame, whereas the direction of this difference was reversed when the base rate was high. Importantly, the findings also show that in the intermediate base-rate condition ($p = 0.5$), in which the evidence should be about equally compatible with the occurrence or the non-occurrence of success, the simple effect of frame on probability judgment was nonsignificant – in fact, the means are identical. Thus, as RAP predicts, the strict refocusing effect was eliminated when the available evidence was roughly as compatible with success as with failure.

In support of the predicted pessimistic bias, but inconsistent with the predictions of MMT's proportionality principle, probability estimates in the 0.5 condition varied as a function of the event frequency × frame interaction, $F(2,42) = 7.41$, $MSE = 0.06$, $P < 0.01$. Looking within the rows of Table 8, it is evident that, on average, participants' estimates declined with a three-fold increase in the frequency

Table 7
Mean probability judgments (0–1) in Experiment 5 as a function of base rate and frame

Frame	Base rate			*M*
	0.1	0.5	0.9	
Occurrence	0.10	0.33	0.66	0.35
Non-occurrence	0.66	0.33	0.22	0.40
M	0.35	0.33	0.44	0.37

Table 8
Mean probability judgments (0–1) in Experiment 5 as a function of event frequency and frame in the 0.5 condition

Frame	Event frequency		*M*
	Once	Three times	
Occurrence	0.46	0.21	0.33
Non-occurrence	0.27	0.39	0.33
M	0.36	0.30	0.33

of success (i.e., landing on a winning strip), whereas their estimates increased with a three-fold rise in the frequency of failure (i.e., not landing on a winning strip). Examining the data within columns of Table 8, it is evident that participants judged a single success out of four trials as almost twice as likely as a single (equiprobable) failure out of four trials, whereas they judged three successes out of four trials as roughly half as likely as three (equiprobable) failures out of four trials. In other words, the simple effects within event frequency also reveal a pessimistic bias because in both cases the more favorable compound event is judged to be less probable. Overall, then, the findings support the hypothesis that people tend toward pessimistic predictions and demonstrate a new form of systematic incoherence in probability judgment – one akin to that observed in earlier studies of surprise (Teigen & Keren, 2002, 2003).

2.6. Experiment 6

Experiments 1–5 demonstrated that subjective probability is influenced by the interaction between how focal events are framed and the base-rate evidence provided. In each experiment, descriptions that were more compatible with the base-rate evidence than their complementary descriptions were assigned significantly greater probability. The objective of Experiment 6 was to examine how coherently participants judge the probability of *distinct* monetary outcomes that were nevertheless equiprobable. Participants were asked to judge the probability of winning \$0, \$10, \$100, \$500, or \$1,000 in the *success* version of the game-show problem. These outcomes correspond to exactly 0–4 successes, respectively, and thus exhaustively vary event frequency. By also manipulating base-rate evidence as in Experiment 5, several different, but equiprobable, outcomes could be compared. For instance, the probability of winning \$500 if $p = 0.1$, $\Phi(\$500|p = 0.1)$, equals that of winning \$10 if $p = 0.9$, $\Phi(\$10|p = 0.9)$.

As noted earlier, it is proposed that the strict refocusing effect results from assessments of the degree of compatibility between representations of focal events and evidence. Experiment 6 tested whether such compatibility assessments are sensitive to category-boundary effects. Kahneman and Tversky (1979) coined the term *category-boundary effect* to refer to the tendency revealed through people's judgments and choices to parse continuous scales of measurement into discrete categories. For instance, event probabilities tend to be parsed into those that are either certain

($P = 1$) or uncertain ($P < 1$), or into those that are either possible ($P > 0$) or impossible ($P = 0$). It is proposed here that a category boundary also exists between the concepts of *something* and *nothing*. In the game-show problem, this would be manifested in terms of the distinction between winning something and winning nothing. Accordingly, it was hypothesized that mathematically equiprobable outcomes that were described in terms of winning positive monetary sums – *any* positive sum – would be judged more probable in the highly compatible, high base-rate condition than in the less compatible, low base-rate condition. Specifically, Inequalities 5–7 in equiprobable monetary outcomes were predicted:

$$\overline{P}(\$10|p = 0.1) < \overline{P}(\$500|p = 0.9) \tag{5}$$

$$\overline{P}(\$100|p = 0.1) < \overline{P}(\$100|p = 0.9) \tag{6}$$

$$\overline{P}(\$500|p = 0.1) < \overline{P}(\$10|p = 0.9). \tag{7}$$

Experiment 6 also provided further tests of the pessimistic bias. It was hypothesized that probability judgments of success would be inversely related to event frequency when success and failure on a given trial were equiprobable. Specifically, Inequalities 8–9 were predicted:

$$\overline{P}(\$0|p = 0.5) > \overline{P}(\$1,000|p = 0.5) \tag{8}$$

$$\overline{P}(\$10|p = 0.5) > \overline{P}(\$500|p = 0.5). \tag{9}$$

The pessimistic bias also entails Inequalities 10–11, which involve equiprobable outcomes in which the base rates of success and failure are not equivalent:

$$\overline{P}(\$0|p = 0.1) > \overline{P}(\$1,000|p = 0.9) \tag{10}$$

$$\overline{P}(\$1,000|p = 0.1) < \overline{P}(\$0|p = 0.9). \tag{11}$$

Inequality 10 captures the prediction that winning nothing when the chances of winning are low is viewed as more likely than winning everything when the chances of winning are correspondingly high. Inequality 11 captures the related prediction that winning everything when the chances of winning are low is viewed as less likely than winning nothing when the chances of winning are correspondingly high.

2.6.1. Method

2.6.1.1. Participants. Participants were 264 students from University of Victoria recruited in the same manner as in Experiment 1.

2.6.1.2. Design and procedure. Participants were randomly assigned to one of twelve experimental conditions in a 3 (Base rate: 0.1, 0.5, 0.9) $\times$ 4 (Event frequency: 1, 2, 3, 4) factorial design. Participants were first presented with the *success* version of the game-show problem described in Experiment 3, and base rate was manipulated between subjects in the scenario as in Experiment 5. Next, depending on the event-frequency condition to which participants were assigned, they were asked:

What is the probability that you will win ($10, $100, $500, $1,000) in this game (i.e., in 4 spins of the wheel)?

A sub-sample of 144 participants was then asked the following two questions:

What is the probability that you will win something in this game? [$k \geqslant 1$]
What is the probability that you won't win anything in this game? [$k = 0$]

The remaining 120 participants were asked only one of these latter two questions on the basis of random assignment. Participants responded to the probability judgment tasks by providing a decimal estimate following the instructions described in Experiment 4.

2.6.2. Results

Preliminary analyses revealed that the mean probability estimates of winning something and of not winning anything did not significantly differ as a function of whether participants were asked to judge both of the probabilities or only one of them (in both cases, $t < 1$). Thus, the data are collapsed across this control factor in subsequent analyses.

Fig. 3 shows the mathematical and mean subjective probabilities as a function of base rate and event frequency. Event frequencies were varied between subjects and the event frequency of 0 was based on the mean response to the "won't win anything" question. As Fig. 3 shows, mean probabilities for $k \geqslant 1$ increased as a monotonic function of base rate ($F(2,252) = 98.75$, $MSE = 0.05$, $P < 0.001$) and decreased as a function of event frequency ($F(2,252) = 9.11$, $P < 0.001$). The interaction effect was nonsignificant, $P = 0.15$. It is evident from Fig. 3 that probability estimates were least sensitive to variations in event frequency in the high base-rate condition, and bias (i.e., systematic inaccuracy; see Ronis & Yates, 1987) was greatest in that condition. Indeed, mean probabilities in the low base-rate condition showed little bias,

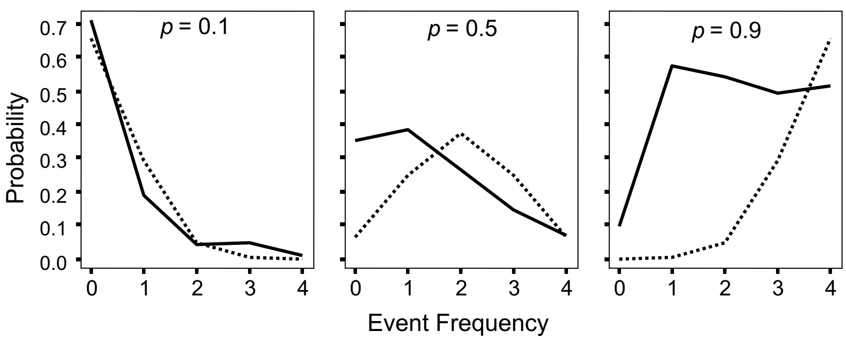

Fig. 3. Mathematical probability (dotted lines) and mean subjective probability (solid lines) in Experiment 6 as a function of base rate and event frequency.

as did mean probabilities in the intermediate base-rate condition for event frequencies greater than one.

The pattern of results observed in Fig. 3 also reveals strong subadditivity in the high base-rate condition. As Table 9 shows, packed estimates of mean total probability, $\overline{P}(\$0) + \overline{P}$ (\$10 or more), did not differ significantly as a function of base rate; and, consistent with past research (e.g., Idson et al., 2001; Macchi et al., 1999; Mandel, 2005) the grand mean was slightly, but reliably, superadditive, one-sample $t(143) = -2.55$, $P < 0.02$. By contrast, the unpacked estimates of mean total probability, $\overline{P}(\$0) + \overline{P}(\$10) + \overline{P}(\$100) + \overline{P}(\$500) + \overline{P}(\$1,000)$, increased as a positive monotonic function of base rate. Thus, the unpacking factor (i.e., the unpacked estimate divided by the packed estimate; see Tversky & Koehler, 1994) also increased with base rate, in line with the *enhancement effect* – the tendency for the sum of the probabilities assigned to exhaustive and mutually-exclusive partitions of a set to exceed the probability assigned to the set as a whole when support for each partition is enhanced (Koehler et al., 1997).

The mean probability estimates for the seven contrasts of equiprobable outcomes in Inequalities 5–11 are shown in Table 10. Nine of the 14 mean probabilities differed significantly from the mathematical probability. Each of the five mean probabilities in the high base-rate condition differed significantly from the mathematical probability, whereas only one of the five mean probabilities in the low base-rate condition was significantly inaccurate. More importantly, Table 10 shows that the mean probability estimates for each of the seven contrasts differed significantly in the predicted

Table 9
Mean total probability in Experiment 6 as a function of estimate and base rate

Estimate	Base rate			$F_{(1,261)}$	P
	0.1	0.5	0.9		
Packed	0.93	0.96	0.99	1.02	0.359
Unpacked	0.93	1.23	2.30	506.10	0.000
Unpacking factor	1.00	1.28	2.32		

Table 10
Mean probability judgments in Experiment 6 for seven contrasts of equiprobable compound events

No.	$p_1, k_1: p_2, k_2$	Φ	$\overline{P}_1$	$\overline{P}_2$	$\Delta\overline{P}$	t	P
1	0.1, 0: 0.9, 4	0.6561	0.7079	0.5124*	0.1955	2.79	0.006
2	0.1, 1: 0.9, 3	0.2916	0.1880*	0.4942*	−0.3062	3.67	0.001
3	0.1, 2: 0.9, 2	0.0486	0.0400	0.5414*	−0.5014	8.12	0.000
4	0.1, 3: 0.9, 1	0.0036	0.0484	0.5759*	−0.5275	5.62	0.000
5	0.1, 4: 0.9, 0	0.0001	0.0065	0.0963*	−0.0898	6.78	0.000
6	0.5, 0: 0.5, 4	0.0625	0.3484*	0.0706	0.2778	7.41	0.000
7	0.5, 1: 0.5, 3	0.2500	0.3858*	0.1461*	0.2397	4.85	0.000

Note. p_i = base rate of success per trial, k_i = number of successes in four trials, Φ = mathematical probability, $\overline{P}_i$ = mean subjective probability, $\Delta\overline{P} = \overline{P}_1 - \overline{P}_2$, t is a test of the difference between $\overline{P}_1$ and $\overline{P}_2$. $*\overline{P}_i \neq \Phi$, $P < 0.03$.

direction. In particular, the results of contrasts 2–4 provide support for Inequalities 5–7, respectively. These findings support the hypothesis that subjective probability is influenced by compatibility between representations of the focal event and evidence, and that outcomes involving winning are grouped into the ad-hoc category of "winning something." The results of contrasts 1 and 5–7 each demonstrate the pessimistic bias. Respectively, contrasts 1 and 5 provide support for Inequalities 10 and 11, whereas contrasts 1 and 5 provide support for Inequalities 8 and 9. These findings replicate the pessimistic bias demonstrated in Experiment 5 with outcomes described in monetary terms and further show that the pessimistic bias is evident in cases in which the base rates on which equiprobable outcomes are conditioned are themselves unequal.

3. General discussion

The present research documented two novel violations of coherence in subjective probability. First, Experiments 1–5 revealed that judgments of the probability of compound events were influenced by the interaction of strict refocusing and evidence. Specifically, compared to compound events framed in terms of k occurrences out of n trials, the same events framed in terms of the complementary number of non-occurrences were judged (a) less probable when the base rates favored event occurrence, (b) more probable when the base rates favored event non-occurrence, and (c) equally probable when the base rates indicated the equiprobability of event occurrence and non-occurrence. These findings were evident despite variations in the problem content, the wording of the probability judgment task, and the response mode, and despite the fact that the mathematical probability was deducible from the information provided. Second, Experiments 5 and 6 revealed a pessimistic bias in which the less favorable of two equiprobable compound events was, on average, judged as the more probable one. This effect was evident regardless of whether the compound events were described in terms of the number of conjoint occurrences, the number of conjoint non-occurrences, or their monetary value.

3.1. Moderation of strict refocusing effects by evidence as a test of RAP

In the present article, these novel findings were situated within a theoretical account (RAP) that addresses how representational and assessment processes interactively determine probability judgment. Two principles – focalism and redundancy reduction – were posited to account for important aspects of representation in probability judgment. Like MMT, RAP proposes that focal events are represented in terms of features that are explicit in their description. The emphasis in RAP, however, is on the psychological property of focalism, whereas in MMT it is on logical property of truth. In the present context, the focalism principle implies that strict refocusing will systematically alter the representation of events.

RAP permits representations to correspond to logical possibilities as in MMT, but, unlike MMT, RAP does not preclude other representations that may more parsimoniously encode relevant information in a given context. In the present research, the two accounts differ in terms of how the event-generating model and base-rate evidence are represented. Based on the redundancy-reduction principle, RAP proposes a parsimonious distributional representation of outcomes on a given trial as depicted on the right side of Fig. 1. This representation applies equally well to any trial length and, thus, does not increase demands on working memory as trial length increases. MMT, by contrast, precludes this type of representation because it does not correspond to the theory's basic assumption that "each possibility that [people] envisage is represented in a separate mental model, which as far as possible has an iconic structure" (Johnson-Laird, 2006, p. 123; see Over, 2005, for further discussion of this limitation). This constraint of MMT is psychologically implausible for problems involving large numbers of true possibilities because representing each one would exceed the capacity of working memory. It is therefore not surprising that tests of MMT have tended to rely on problems entailing relatively small numbers of true possibilities (e.g., see Girotto & Gonzalez, 2005; Johnson-Laird et al., 1999).

According to RAP, representations provide the direct input into the assessment process. In the present context, the compatibility principle proposes that the representation of a focal event is assessed in terms of its perceived compatibility with the distributional representation of the event-generating model. Based on the three principles – focalism, redundancy-reduction, and compatibility – RAP predicts that evidence indicative of event occurrence will be most compatible with an occurrence frame, evidence indicative of event non-occurrence will be most compatible with a non-occurrence frame, and evidence equally indicative of event occurrence or non-occurrence will be about equally compatible with either frame. This is precisely what the cross-over interaction effects between frame and base-rate evidence revealed in Experiments 1–5. Moreover, Experiment 4 directly demonstrated the mediating effect of perceived compatibility, and Experiment 5 demonstrated the elimination of the strict refocusing effect when success and failure were equally compatible with the base-rate evidence. Experiment 6 further demonstrated that probability judgments of monetary outcomes conformed to the compatibility principle such that equiprobable outcomes in the ad-hoc category of "winning something" were assigned greater probability if they were conditional on a high rather than low base rate of success. The latter findings support a key proposition of judgment dissociation theory (Mandel, 2003), which states that ad hoc categories play an important role in defining the meaning (and, thus, the extension) of the event being judged.

Although the present findings supported the compatibility principle, an alternative explanation in terms of anchoring and adjustment (Tversky & Kahneman, 1974) must be considered. According to the anchoring account, participants who judged $P(X = 1| p = 0.9)$, for example, would anchor their estimate at 0.9 and then adjust their estimate downward in light of the three non-occurrences. Given that adjustment is often insufficient, this process could explain why participants overestimated this probability. However, the anchoring account has difficulty explaining the estimates of participants who judged $P(\neg X = 3| p = 0.9)$ because they would be expected to anchor their estimate at 0.1 (for 1 non-occurrence) and then adjust downward in light of the remaining two non-occurrences and one occurrence. In contrast, mean probabilities in that condition were always greater than 0.1. This would require not only insufficient adjustment, but adjustment in the wrong direction, which is inconsistent with the account. As well, the account does not easily accommodate the fact that perceived representativeness mediated the interactive effect of frame and base rate on probability judgments in Experiment 4. Finally, according to the anchoring account, mean probabilities of $X = 1$, $X = 2$, $X = 3$, and $X = 4$ in the high base-rate condition of Experiment 6 ought to be sensitive to event frequency, whereas in fact they were not. Overall, then, the findings more strongly favor a compatibility-matching process.

3.2. The pessimistic bias and compatibility with belief

The pessimistic bias observed in this research suggests that probability judgments are also influenced by the compatibility of focal events and prior beliefs about the distribution of event types in everyday life. It was proposed that, in daily life, people learn that uncontrollable event-generating processes are more likely to yield negative outcomes than positive ones. Supporting this hypothesis, but contrary to the predictions of MMT, participants in Experiment 5 judged multiple successes as less probable than a single, equiprobable success, whereas they judged multiple failures as more probable than a single, equiprobable failure. Moreover, participants in Experiment 6 assigned greater probability to winning the worst possible outcome than to winning the best possible outcome in the game-show problem even though the mathematical probabilities of these monetary outcomes were identical.

This evidence of pessimistic bias in probability judgment builds on Teigen and Keren's (2002, 2003) findings showing that uncontrollable successes are more surprising than equally uncontrollable and equiprobable failures. A key difference between the two sets of findings, however, is that whereas the assignment of differing degrees of probability to equiprobable events clearly violates the normative principle of extensionality, the elicitation of differing degrees of surprise by equiprobable events, strictly speaking, cannot be deemed irrational because a normative model of emotional response currently does not exist. In spite of this difference, both sets of findings suggest that, at least in the domain of uncontrollable events, people believe in a variant of Murphy's Law (i.e., *if something bad can happen, it will*) – namely, that among uncontrollable equiprobable events, the least favorable one is seen as more likely to occur. The basis of this belief is currently unknown. One possibility is that it reflects accurate learning and that the hedonic positivity of uncontrollable outcomes is a positively-skewed distribution. Another possibility is that because people tend to be loss averse (Kahneman & Tversky, 1979), they attend to, and subsequent recall, negative events with greater ease. This appears to be part of the reason why students are reluctant to change their answers on multiple-choice tests. They recall the times they switched from a correct answer to a wrong one much more readily than the times they switched from a wrong answer to a correct one, with this negative-recall bias being predictive of reluctance to switch (Kruger, Wirtz, & Miller, 2005). At face value, the pessimistic bias may seem to be inconsistent with the preferences of many to purchase lotteries having exceedingly small chances of yielding winning outcomes, especially large jackpots. However, as behavioral decision research has shown, people tend to either overweight or ignore small probabilities (Kahneman & Tversky, 1979). The decision to purchase lotteries likely reflects such overweighting in decision making and is in fact not incompatible with the pessimistic bias.

At face value, the pessimistic bias may also appear to contradict the literature showing that people tend to be *unrealistically optimistic* in comparing themselves to others (e.g., Armor & Taylor, 2002; Dhami, Mandel, Loewenstein, & Ayton, 2006; Weinstein, 1980). For instance, *most* people tend to think they are doing *better than average*. The contradiction between unrealistic optimism and the pessimistic bias, however, is more apparent than real. First, whereas findings of unrealistic optimism are based on comparative judgments about the self, the pessimistic bias observed in this research is based on non-comparative judgments about possible states of the world. Second, unrealistic optimism is evident in domains in which outcomes are at least somewhat controllable. People do not judge their chances of uncontrollable events such as winning the lottery as being better than others' chances (Weinstein, 1980; Zakay, 1984). In contrast, the pessimistic bias has thus far only been demonstrated in connection with uncontrollable events. Indeed, Teigen and Keren (2002) found that, among equiprobable controllable events, negative outcomes elicit greater surprise than positive ones. Future research could examine whether the pessimistic bias is evident in relative probability judgment. For instance, if participants presented with pairs of equiprobable events differing in favorability were asked to indicate which one was more likely to occur, would they judge the less-favorable event to be more probable, or would they regard them as equally probable? Future research could also test the hypothesis that the pessimistic bias would be eliminated or even reversed in domains in which the outcomes were perceived as controllable.

4. Conclusion

The present account proposes that subjective probability is determined by the interaction of representational and assessment processes. At the representational level, event features are mapped onto cognitive representations, either via direct perception of the events or (as in the present research) via the intermediate stage of event description. How events are described will influence representation by making explicit certain exemplars (as in unpacking) or features (as in strict refocusing) to the exclusion of others, and representations will favor parsimony through the avoidance of redundancy and other simplifying processes. At the assessment level, the compatibility of these representations to representations of evidence is gauged, with subjective probability being a direct function of this assessment. This article sketched some possible ways that framing might influence representation and thus impact the assessment process. These predictions were strongly supported and, overall, RAP provided a more comprehensive account of the findings than MMT, despite their common emphasis on parsimony of mental representation. Future research could profitably refine our understanding of the representational and assessment processes underlying human judgment and provide insights into why judgments often deviate systematically from coherence criteria that represent hallmarks of rational thought.

Acknowledgements

This research was funded by Discovery Grant #249537-2002 from the Natural Sciences and Engineering Research Council of Canada. Portions of this research were presented in May 2006 at the Southern Ontario Behavioral Decision Research Conference in Toronto, Canada and in June 2006 at the Annual Conference of the Institute for Operations Research and the Management Sciences in Hong Kong, China. I thank Derek Koehler, Jacques Mehler, and three anonymous reviewers for their feedback on this research.

References

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85*, 249–277.

Armor, D. A., & Taylor, S. E. (2002). When predictions fail: The dilemma of unrealistic optimism. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 334–347). Cambridge, U.K: Cambridge University Press.

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

Braine, M. D. S., & O'Brien, D. P. (1991). A theory of if: a lexical entry, reasoning program, and pragmatic principles. *Psychological Review, 98*, 182–203.

Dhami, M. K., Mandel, D. R., Loewenstein, G., & Ayton, P. (2006). Prisoners' positive illusions of their post-release success. *Law and Human Behavior, 30*, 631–647.

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

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

Girotto, V., & Gonzalez, M. (2005). Probabilistic reasoning and combinatorial analysis. In V. Girotto & P. N. Johnson-Laird (Eds.), *The shape of reason: Essays in honour of Paolo Legrenzi* (pp. 161–175). New York: Psychology Press.

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* (Second ed., pp. 651–676). Cambridge, U.K: Cambridge University Press.

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.

Johnson-Laird, P. N. (2006). Models and heterogeneous reasoning. *Journal of Experimental and Theoretical Artificial Intelligence, 18*, 121–148.

Johnson-Laird, P. N., Byrne, R. M. J., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review, 99*, 418–439.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., & Legrenzi, M. S. (2000). Illusions in reasoning about consistency. *Science, 288*, 531–532.

Johnson-Laird, P. N., Legrenzi, P., Girotto, V., Legrenzi, M. S., & Caverni, J.-P. (1999). Naive probability: a mental model theory of extensional reasoning. *Psychological Review, 106*, 62–88.

Johnson-Laird, P. N., & Savary, F. (1999). Illusory inferences: a novel class of erroneous deductions. *Cognition, 71*, 191–229.

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

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology, 3*, 430–454.

Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica, 47*, 263–291.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201–208). New York: Cambridge University Press.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology (Fourth ed.. In D. Gilbert, S. Fiske, & G. Lindzey (Eds.). *The handbook of social psychology* (Vol. 1, pp. 233–265). Boston, MA: McGraw-Hill.

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

Kruger, J., Wirtz, D., & Miller, D. T. (2005). Counterfactual thinking and the first instinct fallacy. *Journal of Personality and Social Psychology, 88*, 725–735.

Legrenzi, P., Girotto, V., & Johnson-Laird, P. N. (1993). Focussing in reasoning and decision making. *Cognition, 49*, 37–66.

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

Mandel, D. R. (2001). Gain-loss framing and choice: separating outcome formulations from descriptor formulations. *Organizational Behavior and Human Decision Processes, 85*, 56–76.

Mandel, D. R. (2003). Judgment dissociation theory: an analysis of differences in causal, counterfactual, and covariational reasoning. *Journal of Experimental Psychology: General, 137*, 419–434.

Mandel, D. R. (2005). Are risk assessments of a terrorist attack coherent?. *Journal of Experimental Psychology: Applied 11*, 277–288.

Miller, D. T., Turnbull, W., & McFarland, C. (1989). When a coincidence is suspicious: the role of mental simulation. *Journal of Personality and Social Psychology, 57*, 581–589.

Moldoveanu, M., & Langer, E. (2002). False memories of the future: a critique of the applications of probabilistic reasoning to the study of cognitive processes. *Psychological Review, 109*, 358–375.

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

Over, D. (2005). Naïve probability and its model theory. In V. Girotto & P. N. Johnson-Laird (Eds.), *The shape of reason: Essays in honour of Paolo Legrenzi* (pp. 139–160). New York: Psychology Press.

Pacini, R., & Epstein, S. (1999). The relation between rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of Personality and Social Psychology, 76*, 972–987.

Richardson, J., & Ormerod, T. C. (1997). Rephrasing between disjunctives and conditionals: mental models and the effects of thematic content. *The Quarterly Journal of Experimental Psychology, 50A*, 358–385.

Ronis, D. L., & Yates, J. F. (1987). Components of probability judgment accuracy: individual consistency and effects of subject matter and assessment method. *Organizational Behavior and Human Decision Processes, 40*, 193–218.

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

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

Schkade, D. A., & Kahneman, D. (1998). Does living in California make people happy? A focusing illusion in judgments of life satisfaction. *Psychological Science, 9*, 340–346.

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.

Sloutsky, V. M., & Goldvarg, Y. (2002). Mental representation of logical connectives. *The Quarterly Journal of Experimental Psychology, 57A*, 636–665.

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.

Smith, E. E., Shafir, E., & Osherson, D. (1993). Similarity, plausibility, and judgments of probability. *Cognition, 49*, 67–96.

Teigen, K. H. (1983). Studies in subjective probability III: the unimportance of alternatives. *Scandinavian Journal of Psychology, 24*, 97–105.

Teigen, K. H. (2001). When equal chances = good chances: verbal probabilities and the equiprobability effect. *Organizational Behavior and Human Decision Processes, 85*, 77–108.

Teigen, K. H., & Keren, G. (2002). When are successes more surprising than failures? *Cognition and Emotion, 16*, 245–268.

Teigen, K. H., & Keren, G. (2003). Surprises: low probabilities or high contrasts?. *Cognition 87*, 55–71.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185*, 1124–1130.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science, 211*, 453–458.

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

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology, 39*, 806–820.

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.

Wright, G., & Whalley, P. (1983). The supra-additivity of subjective probability. In B. P. Stigum & F. Wenstop (Eds.), *Foundations of utility and risk theory with applications* (pp. 233–244). Dordrecht, The Netherlands: Reidel.

Zakay, D. (1984). The influence of perceived event's controllability upon its subjective occurrence probability. *Psychological Record, 34*, 233–240.