Title
Do Evaluation Frames Improve the Quality of Conditional Probability Judgment?

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 29(29)

ISSN
1069-7977

Authors
Williams, Joseph Jay
Mandel, David R.

Publication Date
2007

Peer reviewed

Do Evaluation Frames Improve the Quality of Conditional Probability Judgment?

Joseph Jay Williams (joseph.williams@utoronto.ca) and David R. Mandel (mandel@psych.utoronto.ca)
Department of Psychology, University of Toronto, Toronto, ON, M5S 3G3, Canada

Abstract

In *evaluation frames*, both focal and alternative hypotheses are explicit in queries about an event's probability. We investigated whether evaluation frames improved the accuracy and coherence of conditional probability judgments when compared to *economy frames* in which only the focal hypothesis was explicit. Participants were presented with contingency information regarding the relation between viruses and an illness with an unknown etiology, and they judged the conditional probability that the illness would occur or not occur given that a virus was either present or absent. Compared to economy frames, evaluation frames improved the accuracy and coherence of probability judgments.

Keywords: probability; judgment; evaluation frame; coherence; accuracy; hypothesis testing

Introduction

Judgments of probability represent a fundamental aspect of human cognition. Many everyday and professional situations require judging the likelihood of an event: What are the odds that it will rain tomorrow? How likely are stocks to fall this week? There are also many situations in which the implicit evaluation of an event's likelihood determines what actions are taken, such as purchasing one life insurance plan over another (Johnson, Hershey, Meszaros, & Kunreuther, 1993). In this paper, we propose and test the efficacy of one general method for improving probability judgment.

The present research is motivated by a key insight of Tversky & Koehler's (1994) *support theory*, which offers a non-extensional account of probability judgment. Support theory posits that people assign probabilities to *descriptions* of events (called *hypotheses*) rather than to the events *per se*. Support theory also posits that probability judgments are assessed using an *evaluation frame* (F, A), where $P(F \text{ v. } A)$ is the probability that the focal hypothesis, F, is true rather than the mutually-exclusive alternative hypothesis, A. For example, consider the query: "How likely is it to rain tomorrow rather than be sunny?" Support theory assumes that this query is represented in an evaluation frame in which F is "it rains tomorrow" and A is "it is sunny tomorrow." According to the theory, probability judgment involves assessing the evidential support for F over the total support for F and A.

In much of everyday conversation, however, interlocutors are unlikely to pose queries to each other in evaluation frames because of the conversationally-normative injunction to be brief and to the point (Grice, 1975). This conversational norm dictates that queries will often leave the alternative hypothesis implicit. For example, the query "How likely is it to rain tomorrow?" makes the focal hypothesis explicit, but does not specify its alternative. We refer to queries posed in this manner as *economy* frames to indicate that they are economical in providing just enough information to specify the event of primary interest.

A key question addressed in the present research was whether economy of expression yields a cost in terms of the resultant quality of probability judgments. To examine this question, we posed identical sets of probability queries to two groups of participants, both of which were presented with identical information pertinent to their assessments. The only difference was that one group was queried using evaluation frames and the other using economy frames.

Although evaluation framing is regarded as a descriptive concept in support theory, we hypothesized that it may also be of prescriptive importance as a method for improving the quality of probability judgment, especially the *coherence*, or logical consistency, of judgment. In particular, we hypothesized that evaluation frames would increase the coherence of judgment in cases in which the alternative to a focal hypothesis F is the mutually exclusive and exhaustive set, $A = \neg F$. Moreover, we argue, it is precisely in such cases that queries about probability are most likely to posed in economy frames. This is because the alternative hypothesis constitutes the default, logical complement of the focal hypothesis. Therefore, it is directly inferable from F.

In probability calculus, the logical relation between the probabilities of complements is captured by 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 any one of the events will occur (Edwards, 1982). When the events in question are binary complements, such as "Liberals win the election" and "Liberals lose the election," the additivity property is given a straightforward interpretation: these events must sum to unity—namely, 1.0 on the [0, 1] probability scale (or 100 on its percentage equivalent).

Although early studies of the coherence of probability judgments of binary complements found that, on average, people's assessments were additive (e.g., Wallsten, Budescu, & Zwick, 1992; Tversky & Koehler, 1994), a number of subsequent studies report superadditivity in such judgments (e.g., Macchi, Osherson, & Krantz, 1999; Mandel, 2005)—namely, that the sum of the complementary probabilities is significantly less than unity.

In an attempt to resolve this inconsistency in the literature, Mandel (2005) noted that studies yielding evidence of additive judgments tended to query participants using evaluation frames, whereas studies yielding evidence of superadditivity tended to use economy frames instead.

Although suggestive, Mandel (2005) did not put this potential moderator to an empirical test. To our knowledge, this paper reports the first direct test of this hypothesis.

There are a number of reasons why evaluation frames might improve the coherence of probability judgments. First, many studies reveal that the mere act of imagining a possibility as true, such as the event specified in a probability query, increases its perceived likelihood (e.g., see Koehler, 1991). This is consistent with experimental evidence supporting the Spinozan view that communicated propositions are initially coded as true and only later verified if the individual is not cognitively overtaxed (Gilbert, 1991). By making alternative and focal hypotheses equally explicit, evaluation frames may attenuate, if not eliminate, the effect of this truth bias on judgment.

Second, people have a tendency to assess hypotheses using what Klayman and Ha (1987) called a *positive test strategy*, which involves testing hypotheses by examining cases that conform to the events predicted on the basis of the focal hypothesis. For example, Mandel and Vartanian (2006) found that people presented with 2 × 2 contingency information in a causal judgment task gave the greatest weight both in causal judgment and in ratings of cell importance to the cells implicated in positive testing, and especially to the positive-test cell that confirmed the focal hypothesis. In other words, the weighting of information was biased toward the cells specified in the focal hypothesis, and especially the cell that supported it. By presenting hypotheses in evaluation frames, this bias ought to be eliminated because competing hypotheses are both explicit, thus canceling the effect of positive testing, and perhaps minimizing the effect of confirmation bias.

Finally, evaluation frames can serve to remind people of the requirement that complementary probabilities must sum to unity. That is, logically speaking, when people assign a probability to an event, x, it is implied that the probability assigned to x's complement, $\neg x$, is $1 - P(x)$. People, however, often assign probabilities to binary complements such that they do not add up to 1. Evaluation frames may give people more insight into the nature of this logical constraint by making the complementarity relation more salient. In line with the suggestion that explicit awareness of the complementarily of two judgments can improve coherence, Mandel (2005) found that superadditivity was attenuated when complementary probability judgments were made in close enough proximity that their relationship was transparent. This suggests that violations of coherence represent *application errors*—namely, failures to apply relevant logical principles—rather than comprehension errors—namely, failures to understand such principles (Kahneman & Tversky, 1982). We posit that when focal and alternative hypotheses represent binary complements, evaluation frames may be one method for highlighting their logical relation and the additivity requirement.

Experiment

The present experiment investigated the effect of framing on the quality of conditional probability judgments. To do this, we used a trial-by-trial learning paradigm common in many studies of causal induction and contingency judgment (e.g., Mandel & Lehman, 1998). By controlling the information on which participants were asked to base their assessments, we were able to examine not only the coherence of their responses, but also their accuracy. Moreover, this degree of control permitted us to examine how our manipulation of framing might interact with the mathematical probability of the focal hypothesis and its alternative to influence coherence and accuracy.

After the probability judgment phase of the experiment, we also asked participants to rate the importance of the four contingency cells in answering each type of query posed. This allowed us to test the hypothesis that evaluation frames would prompt a more equal weighting of relevant sources of information than economy frames, by making the cell presenting hypothesis-disconfirming information more salient through the description of the alternative hypothesis. And, should we find support for that hypothesis, we would be able to test whether the discrepancy in importance assigned to the two relevant cells mediated any observed effect of framing on judgment accuracy.

Method

Participants Participants were 40 University of Toronto undergraduates who received $12.00 for their participation.

Procedure and Design: Part I Participants were initially presented with a booklet that described the experimental scenario. Participants were asked to imagine that they were a military analyst investigating the recent outbreak of a new illness, thought to be the result of a biological terrorist attack. They were informed that the illness was suspected to be caused by a genetically-modified viral agent, although the specific agent had not been identified. Their task was to assess the conditional probability of the illness either occurring or not occurring given that a target viral agent was either present or absent. The evidence on which participants were asked to make these judgments was 2 × 2 contingency information presented on a trial-by-trial basis in which each trial corresponded to 1 of 20 "patient" records.

Participants were asked to make 28 conditional probability judgments in all. Prior to each judgment, participants viewed the "test results" of a sample of 20 patients; that is, participants observed 20 trials in sequence that presented contingency information that could be categorized in terms of the following four cells: (a) virus present and illness present ($V \bullet I$), (b) virus present and illness absent ($V \bullet \neg I$), (c) virus absent and illness present ($\neg V \bullet I$), and (d) virus absent and illness absent ($\neg V \bullet \neg I$). Each patient's test result was displayed onscreen using two circles. The left circle was labeled "Virus" and the right circle was labeled "Illness." Each circle represented the status (i.e., present or absent) of the target virus and the illness in the patient. If the virus or illness was present, the corresponding circle was green and the word "Present" was displayed in the center; when absent, the circle was red and the word “Absent” was displayed. Participants were shown examples of the four types of test results in the introductory booklet. Each of the 20 trials was displayed onscreen for 2.5 s, with a 1.5 s “++” inter-trial mask.

After viewing a trial sequence, participants were asked to answer a particular probability query by using the arrow keys to move a marker along a scale displayed onscreen. The scale ranged from 0.0 to 1.0 with notches labeled at each 0.1 increment. The marker’s default position was at the 0.5 notch and the marker moved in increments of 0.05. For ease of presentation, these values were multiplied by 100 in subsequent analyses.

The experiment used a 2 (Frame) × 4 (Query) × 7 (Distribution) mixed design. Frame was manipulated between subjects, and Query and Probability were manipulated within subjects. Table 1 shows the four queries presented in the evaluation-frame condition. Corresponding queries in the economy-frame condition were identical except that the alternative hypothesis was not described. That is, the expression “rather than ___” was omitted.

Table 1: Queries in the Evaluation-Frame Condition

Query	Wording
$P(I\|V)$	When the virus is present, how likely is the illness to be present rather than absent?
$P(\neg I\|V)$	When the virus is present, how likely is the illness to be absent rather than present?
$P(I\|\neg V)$	When the virus is absent, how likely is the illness to be present rather than absent?
$P(\neg I\|\neg V)$	When the virus is absent, how likely is the illness to be absent rather than present?

The 28 conditional probability judgments each participant provided were obtained by crossing these four types of query by the seven different sample distributions shown in Table 2. The presentation order of these 28 question-distribution conjunctions was randomly generated for each participant, as was the presentation order of the 20 test results in each patient sample.

Table 2: Sample Distributions of Cell Frequencies

	Cell			
Distribution	$V \bullet I$	$V \bullet \neg I$	$\neg V \bullet I$	$\neg V \bullet \neg I$
1	10	0	0	10
2	8	2	2	8
3	6	4	4	6
4	5	5	5	5
5	4	6	6	4
6	2	8	8	2
7	0	10	10	0

Procedure and Design: Part II Following Part 1, participants were shown a 2 × 2 contingency table that depicted the four cells and were reminded that each of the test results they had seen was one of these four types. For each query they answered during the experiment, they were asked to rate how important they thought each type of test result was for answering that query on an 11-point scale ranging from *not at all important* (0) to *absolutely necessary* (10). Thus, participants’ ratings conformed to a 4 (Query) × 4 (Cell) within-subjects design. These ratings were blocked by query, with both query order and cell order randomized for each participant.

Results

Coherence To test our prediction that evaluation frames would lead to more coherent judgments than economy frames, we first computed the summed probabilities for binary complements:

$$T^+ = P(I|V) + P(\neg I|V) \text{ and } T^- = P(I|\neg V) + P(\neg I|\neg V).$$

These values were computed at each level of Distribution and a value of 100 was subtracted from the resulting values. Thus, positive and negative values indicate subadditivity and superadditivity, respectively, with a value of zero indicating additive probability judgments. Although the effect of Distribution *per se* on T values was of little interest, it was possible in this design to further examine whether the additivity of binary complements was moderated by the degree to which the available evidence differentially supported the focal and alternative hypotheses. To compute this measure of Discrepancy, we averaged over levels 1 and 7 (high), 2 and 6 (moderate), and 3 and 5 (low) as shown in Table 2, and kept level 4 (none) separate.

The resulting T values were analyzed in a 2 (Frame) × 2 (T: positive vs. negative conditional event) × 4 (Discrepancy) mixed analysis of variance (ANOVA). As predicted, the main effect of Frame was significant, $F(1, 38) = 4.40$, $MSE = 3017.18$, $p < .05$. Participants were more coherent in the evaluation-frame condition ($M = -4.14$, $T = 95.86$) than in the economy-frame condition ($M = -10.88$, $T = 89.12$). However, even in the former condition, participants’ combined probabilities were, on average, superadditive, one-sample $t(19) = 2.38$, $p < .05$. Thus, evaluation frames attenuated but did not fully eliminate superadditivity in probability judgments of binary complements, consistent with the findings of previous research (for a review, see Mandel, 2005). None of the other effects in the model was significant.

Accuracy Although incoherent probability judgments imply inaccuracy, coherent judgments do not necessarily imply accuracy. Accordingly, we examined an inversely proportional measure of accuracy, *bias*, which refers to the mathematical probability of the focal hypothesis subtracted from a participant’s judged probability, $B = J(F) - P(F)$. By combining levels of Query and Distribution, seven mathematical probabilities with values 0, 20, 40, 50, 60, 80, and 100 were testable. For example, consider a sample in which $V \bullet I = 8$, $V \bullet \neg I = 2$, $\neg V \bullet I = 2$, and $\neg V \bullet \neg I = 8$. If followed by the query “When the virus is present, how likely is the illness to be present?," $P(F)$ = 0.8 or 80 on the transposed scale. We refer to the $P(F)$ factor as Probability.

We conducted a 2 (Frame) × 4 (Query) × 7 (Probability) mixed ANOVA on bias. The main effect of Frame was significant: this result is necessitated in our study given that additivity deviations and bias are perfectly correlated. Judgments were less biased in the evaluation-frame condition than in the economy-frame condition. The analysis also revealed a significant main effect of Probability, $F(6, 228) = 52.21$, $MSE = 558.32$, $p < .001$. No other effect in the model was significant. As shown in Fig. 1, participants overestimated zero probabilities and underestimated probabilities greater than 40, with the degree of underestimation increasing toward 100. This pattern of findings represents a lack of sensitivity to the mathematical probabilities, which were fully derivable from the contingency information provided. Fig. 2 plots mean subjective probability as a function of mathematical probability, illustrating this fact. Queries posed in evaluation frames reduce but do not fully eliminate the bias in participants' probability judgments.

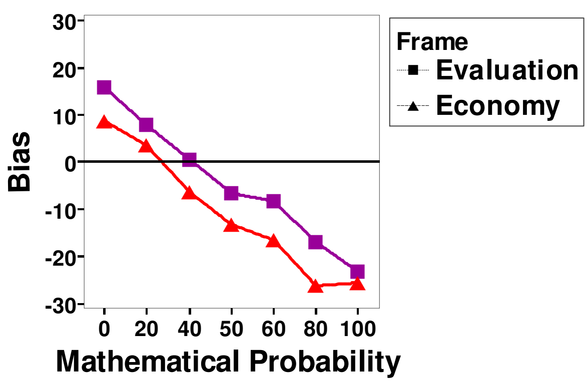

Figure 1: Mean Bias as a Function of Frame and Mathematical Probability.

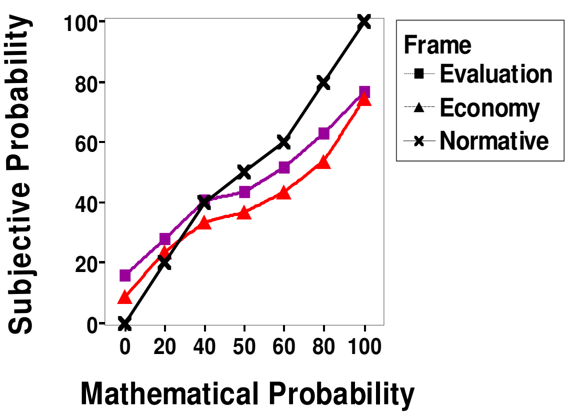

Figure 2: Mean Probability Judgment as a Function of Frame and Mathematical Probability.

Subjective Cell Importance The weighting of contingency information in judgment has been studied most extensively in causal induction tasks (e.g., Mandel & Lehman, 1998; Wasserman, Dorner, & Kao, 1990), where there is some disagreement regarding the normative weighting of the four cells (see, e.g., McKenzie & Mikkelsen, 2007). In contrast, the normative weight of the cells in conditional probability judgment is straightforward.

Each of the four probability queries target a response of the form $P(e|c)$, the probability that event e (I *or* $\neg I$) is true given that condition c (V *or* $\neg V$) is true. Given that $P(e|c) = freq(e \bullet c) \div [freq(e \bullet c) + freq(\neg e \bullet c)]$, $P(e|c)$ completely depends on the cells $(e \bullet c)$ and $(\neg e \bullet c)$ and not at all on the cells $e \bullet \neg c$ and $\neg e \bullet \neg c$. Accordingly, we define the $e \bullet c$ and $\neg e \bullet c$ conjunctions (or cells) as $F^\bullet$ and $A^\bullet$, respectively, to indicate their support of the focal and alternative hypotheses. The two conjunctions in which the necessary condition is false, $e \bullet \neg c$ and $\neg e \bullet \neg c$, are collectively defined as $I^\bullet$ to indicate their irrelevance to the assessment of $P(e|c)$.

We standardized participants' ratings of cell importance by dividing each rating by the sum of the ratings for the four cells, and multiplying this quotient by 10. These standardized weightings were examined as a function of whether the cell supported the focal hypothesis ($F^\bullet$), the alternative hypothesis ($A^\bullet$), or was irrelevant ($I^\bullet$). To examine the degree to which participants were biased in their assessments of cell importance, we subtracted the normative weight (viz., $F^\bullet = A^\bullet = 5$ and $I^\bullet = 0$) from the standardized weight. This yielded a measure of bias for each of the three types of hypothesis-dependent information sources. For example, suppose a participant assessing the importance of the four cells for judging $P(I|V)$ assigned importance ratings of 8, 6, 4, and 2 (on the 0-10 scale) to the $V \bullet I$, $V \bullet \neg I$, $\neg V \bullet I$, and $\neg V \bullet \neg I$ cells, respectively. Their standardized rating would be 4, 3, 2, and 1, and $B_{F^\bullet} = -1$, $B_{A^\bullet} = -2$, and $B_{I^\bullet} = 1.5$, indicating a tendency to underweight $F^\bullet$ and $A^\bullet$ and to overweight $I^\bullet$.

A 2 (Frame) × 3 (Information: $F^\bullet$, $A^\bullet$, $I^\bullet$) ANOVA revealed only a significant main effect of Information on bias in subjective importance, $F(2, 76) = 95.29$, $MSE = 1.48$, $p < .001$. As shown in Fig. 3, this effect is primarily attributable to the predicted overweighting of irrelevant information (and, by implication, the underweighting of relevant information). Fig. 3 also reveals support for the hypothesis that the discrepancy between $B_{F^\bullet}$ and $B_{A^\bullet}$, $\Delta B_{A^\bullet F^\bullet} = B_{A^\bullet} - B_{F^\bullet}$, would be greater in the economy frame than in the evaluation frame. Indeed, the mean discrepancy was significantly greater in the economy-frame condition ($M = -1.37$) than in the evaluation-frame condition ($M = -0.52$), $t(38) = 2.26$, $p < .05$.

Finally, we examined whether our discrepancy measure, $\Delta B_{A^\bullet F^\bullet}$, mediated the predictive effect of frame on bias in probability judgment. As shown in Fig. 4, all conditions for reliable mediation were met: (a) the predictor (Frame) significantly predicted the mediator ($\Delta B_{A^\bullet F^\bullet}$) and the criterion (bias in probability judgment), (b) controlling for the predictor, the mediator significantly predicted the criterion, and (c) the effect of the predictor on the criterion was significantly attenuated after controlling for the mediator, Sobel $t = 2.15$, $p < .05$. Indeed, demonstrating full mediation, Frame was no longer a significant predictor.

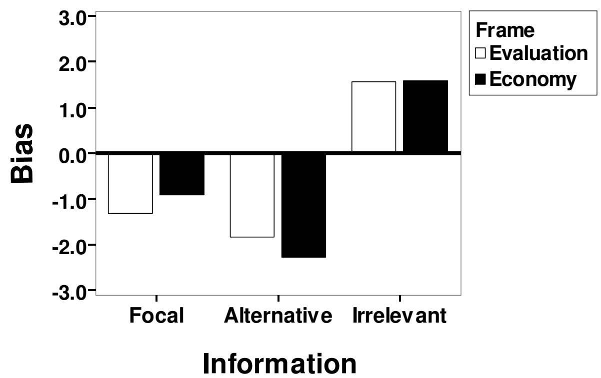

Figure 3: Bias in Information Weighting.

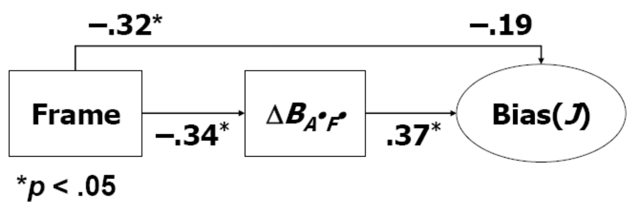

Figure 4: Mediator Model Probability Judgment Bias.

Discussion

The fundamental question that motivated the present research was whether the type of economical descriptive compression common in everyday communication would exact a cost on judgment quality. Specifically, we tested the hypothesis that queries about probability described in terms of evaluation frames would yield more coherent and accurate judgments than the same queries described in terms of economy frames. In support of this hypothesis, participants' probability judgments, on average, were more coherent (i.e., more additive) and accurate (i.e., less biased) when the queries they answered were posed in terms of evaluation frames rather than economy frames.

The findings also shed light on one of the mediating factors that help explain this benefit. First, the findings revealed that evaluation framing led participants to ascribe more equal weight to the two sources of contingency information relevant to the query, in line with normative requirements. Second, the reduction in the discrepancy between the subjective weighting of the two relevant information sources fully mediated the effect of frame on judgment bias. Those findings support the notion that evaluation frames can improve the quality of probability judgments by making the relevance of hypothesis-disconfirming information more salient. In this regard, the findings are consistent with past literature demonstrating that asking people to explain why alternatives to a focal hypothesis might be true tends to debias probability judgments of focal hypotheses (Hirt & Markman, 1995).

However, whereas earlier work suggests that a debiasing effect of evaluation framing might occur by attenuating inflated probability assessments (e.g., Koriat, Fiedler, and Bjork, 2006), we found just the opposite: participants tended to *underestimate* the probability of focal hypotheses. Accordingly, the overall improvement in accuracy was the result of participants assigning *greater* probability to the focal hypothesis in the evaluation-frame condition than in the economy-frame condition. Participants' judgments exhibited *conservatism* (Edwards, 1982)—namely, they overestimated low probabilities and underestimated high probabilities. Thus, the increase in probability associated with evaluation framing had a compensatory effect, reducing bias (on average). Given the counterintuitive nature of this finding, further work is needed to assess the advantage of evaluation framing over economy framing.

In fact, the findings suggest that under certain conditions, it is possible that evaluation frames can lead to *less accurate* judgments. Although the mean differences in subjective probability did not reach statistical significance, this trend is apparent in Fig. 1, when participants were asked to judge events with probabilities of either 20% or 0%. If the present findings prove to be replicable across different judgment tasks, it would suggest that economy frames may be advantageous in domains in which low-probability events require estimation, whereas evaluation frames may be advantageous in domains in which high-probability events require estimation.

More generally, we propose that evaluation frames will aid judgment when focusing on the alternative hypothesis highlights relevant information, and that evaluation frames will impede judgment when doing so obscures relevant information. Such effects, whether salutary or detrimental to judgment quality, lend further support to the *focalism principle*, which states that people tend to represent only those elements of a proposition that are explicit in its description (Mandel, 2007; see also Windschitl, Kruger, & Simms, 2003). An important challenge for future research will be to better understand the conditions under which focalism aids or impedes judgment.

Beyond the effect of framing on judgment quality, two other findings from this experiment deserve brief mention. First, given the cover story that participants read regarding the role of the viruses as potential generative causes of the target illness, one might have anticipated a systematic bias in probability judgments, consistent with that expectation. In this sense, the data were hypothesis-incongruent when $P(I|V)$ was very low and, correspondingly, $P(\neg I|V)$ was very high. Under conditions of high incongruence with prior beliefs, one might expect to observe increased conservatism in judgments. However, the fact that probability judgments were not significantly influenced by the interaction of probability and query type clearly indicates that participants did not display this systematic bias. This may be due to the fact that participants were asked to assess the magnitude of particular conditional probabilities rather than to use those estimates to test the posterior probability of a hypothesis about a particular causal relationship. The absence of a biasing effect of prior beliefs (about the likelihood of a causal relationship) indicates that participants correctly interpreted their task.

Second, whereas many previous studies have focused on the fact that humans tend to overweight hypothesis-confirming information (e.g., Koriat et al., 2006; Mandel & Vartanian, 2006), the present findings revealed that in conditional probability judgments participants assign too much importance to irrelevant information. Understanding why participants overweight irrelevant information in making conditional probability judgments from contingency data is an interesting question for further research.

Conclusion

The present experiment was the first to systematically examine the effect of economy versus evaluation framing on probability judgments based on trial by trial contingency information. On average, evaluation-framed queries prompted judgments that were more coherent and more accurate than those prompted by economy-framed queries. Given the generality of the economy-evaluation frame distinction in expressing hypotheses and queries about them, there is a wide range of judgment tasks in which this issue could be further investigated. Some practical settings in which this issue might be further investigated include risk assessments or analytic forecasts of events occurring rather than not occurring, medical judgments such as assessing a patient's chances of survival rather than mortality, and decision making under risk in which the probability of consequences occurring rather than not occurring is likely to affect choice.

Acknowledgements

This research was funded by Natural Sciences and Engineering Research Council of Canada Discovery Grant #249537-2002 (Mandel). We are grateful to Oshin Vartanian for his valuable input in developing the materials for this research and we thank four anonymous reviewers for their feedback on an early version of this paper.

References

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.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46,* 107-119.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan , *Syntax and semantics Volume 3*: *Speech Acts* (pp. 41-58). New York : Academic Press.

Hirt, E. R., & Markman, K. D. (1995). Multiple explanation: A consider-an-alternative strategy for debiasing judgments. *Journal of Personality and Social Psychology, 69,* 1069-1086.

Johnson, E. J., Hershey, J., Meszaros, J., & Kunreuther, H. (1993). Framing, probability distortions, and insurance decisions. *Journal of Risk and Uncertainty, 7,* 35-51.

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

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94*, 211-228.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin*, *110,* 499-519.

Koriat, A., Fiedler, K., & Bjork, R. A. (2006). The inflation of conditional predictions. *Journal of Experimental Psychology: General, 135,* 429-447.

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

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

Mandel, D. R. (2007). Violations of coherence in subjective probability: A representational and assessment processes account. *Cognition*, doi:10.1016/j.cognition.2007.01.001

Mandel, D. R., & Lehman, D. R. (1998). Integration of contingency information in judgments of cause, covariation, and probability. *Journal of Experimental Psychology: General, 127,* 269-285.

Mandel, D. R., & Vartanian, O. (2006). Is the weighting of contingency data contingent on the hypothesis assessed? In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (p. 2652). Mahwah, NJ: Erlbaum.

McKenzie, C. R. M., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology, 54,* 33-61.

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.

Wasserman, E. A., Dorner, W. W., & Kao, S. F. (1990). The contributions of specific cell information to judgments of interevent contingency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 509-521.

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.