

General Article

DO FREQUENCY REPRESENTATIONS ELIMINATE CONJUNCTION EFFECTS?

An Exercise in Adversarial Collaboration

By Barbara Mellers,¹ Ralph Hertwig,² and Daniel Kahneman³

¹Ohio State University; ²Max Planck Institute for Human Development, Berlin, Germany; and ³Princeton University

The present article offers an approach to scientific debate called adversarial collaboration. The approach requires both parties to agree on empirical tests for resolving a dispute and to conduct these tests with the help of an arbiter. In dispute were Hertwig's claims that frequency formats eliminate conjunction effects and that the conjunction effects previously reported by Kahneman and Tversky occurred because some participants interpreted the word "and" in "bank tellers and feminists" as a union operator. Hertwig proposed two new conjunction phrases, "and are" and "who are," that would eliminate the ambiguity. Kahneman disagreed with Hertwig's predictions for "and are," but agreed with his predictions for "who are." Mellers served as arbiter. Frequency formats by themselves did not eliminate conjunction effects with any of the phrases, but when filler items were removed, conjunction effects disappeared with Hertwig's phrases. Kahneman and Hertwig offer different interpretations of the findings. We discuss the benefits of adversarial collaboration over replies and rejoinders, and present a suggested protocol for adversarial collaboration.

This article presents a new method for resolving scientific debate, which we call adversarial collaboration. The authors did not agree when they began the collaboration, nor do they agree now. But theoretical agreement is not necessary for a successful adversarial collaboration. Success should be gauged by the extent to which joint efforts yield surprising results, insightful discussions, and testable hypotheses about outstanding issues. In Table 1, we provide suggestions for conducting an adversarial collaboration based on a protocol originally developed by Kahneman, but adapted in light of our experiences. For another project in a similar spirit, see Gilovich, Medvec, and Kahneman (1998).

Our starting point is a controversy about biases in probabilistic reasoning in general and conjunction effects in particular (Gigerenzer, 1996; Kahneman & Tversky, 1996). We focus on

a story known as the Linda problem (Tversky & Kahneman, 1982, 1983). Participants were told:

Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Participants then ranked statements about Linda according to their probability. These statements included "Linda is a bank teller," and "Linda is a bank teller and is active in the feminist movement." The majority of respondents judged "Linda is a bank teller and is active in the feminist movement" as more probable than "Linda is a bank teller" in both within-subjects designs (in which participants judged both statements) and between-subjects designs (in which different groups judged the two statements).

Tversky and Kahneman (1983) pointed out that these responses violate the conjunction rule, according to which the probability of the intersection of two events cannot exceed the probability of either single event. Furthermore, they argued that participants based their judgments on the representativeness heuristic: Linda was judged as more likely to be a bank teller and active in the feminist movement than to be a bank teller because she resembles a prototypical feminist bank teller more than a prototypical bank teller. Both claims have sparked considerable controversy (e.g., Adler, 1984; Dulany & Hilton, 1991; Gigerenzer, 1991, 1996; Politzer & Noveck, 1991).

Our collaboration focuses on conjunction effects with frequency formats. Using a within-subjects design, Tversky and Kahneman (1983, p. 309) found that conjunction effects were markedly reduced when information was presented in the form of frequencies. They used a question about health history and age; later research demonstrated reduced conjunction effects with Linda (e.g., Fiedler, 1988). For example, Hertwig and Gigerenzer (1999) gave participants a story about 200 women who fit the description of Linda. Participants were asked questions such as "How

Address correspondence to Barbara Mellers, Department of Psychology, Ohio State University, Columbus, OH 43210; e-mail: mellers.1@osu.edu.

Table 1. *Suggestions for adversarial collaboration*

1. When tempted to write a critique or to run an experimental refutation of a recent publication, consider the possibility of proposing joint research under an agreed protocol. We call the scholars engaged in such an effort participants. If theoretical differences are deep or if there are large differences in experimental routines between the laboratories, consider the possibility of asking a trusted colleague to coordinate the effort, referee disagreements, and collect the data. We call that person an arbiter.
2. Agree on the details of an initial study, designed to subject the opposing claims to an informative empirical test. The participants should seek to identify results that would change their mind, at least to some extent, and should explicitly anticipate their interpretations of outcomes that would be inconsistent with their theoretical expectations. These predictions should be recorded by the arbiter to prevent future disagreements about remembered interpretations.
3. If there are disagreements about unpublished data, a replication that is agreed to by both participants should be included in the initial study.
4. Accept in advance that the initial study will be inconclusive. Allow each side to propose an additional experiment to exploit the fount of hindsight wisdom that commonly becomes available when disliked results are obtained. Additional studies should be planned jointly, with the arbiter resolving disagreements as they occur.
5. Agree in advance to produce an article with all participants as authors. The arbiter can take responsibility for several parts of the article: an introduction to the debate, the report of experimental results, and a statement of agreed-upon conclusions. If significant disagreements remain, the participants should write individual discussions. The length of these discussions should be determined in advance and monitored by the arbiter. An author who has more to say than the arbiter allows should indicate this fact in a footnote and provide readers with a way to obtain the added material.
6. The data should be under the control of the arbiter, who should be free to publish with only one of the original participants if the other refuses to cooperate. Naturally, the circumstances of such an event should be part of the report.
7. All experimentation and writing should be done quickly, within deadlines agreed to in advance. Delay is likely to breed discord.
8. The arbiter should have the casting vote in selecting a venue for publication, and editors should be informed that requests for major revisions are likely to create impossible problems for the participants in the exercise.

many of the 200 women are bank tellers and are active in the feminist movement?" Hertwig and Gigerenzer found that in within-subjects designs with frequencies, conjunction effects disappeared.

Gigerenzer (1991) proposed a general thesis that frequency representations can reduce, and even eliminate, judgmental errors and biases. He and his collaborators developed models that specify when and why frequency and probability judgments are valid or invalid, according to different norms (e.g., Gigerenzer & Hoffrage, 1995; Gigerenzer, Hoffrage, & Kleinbölting, 1991). In a reply to Gigerenzer, Kahneman and Tversky (1996) argued that (a) frequency representations do not eliminate reliance on judgmental heuristics in general, nor representativeness in particular; (b) within-subjects designs may provide cues about the correct rule that enable people to overcome biases—and frequency formats make some cues especially transparent; and (c) between-subjects designs are most appropriate for studying the biases associated with heuristics.

To demonstrate their claims, Kahneman and Tversky (1996) conducted both a within-subjects test and a between-subjects test of conjunction effects. They gave three groups of students the Linda problem in a frequency format. The first group estimated the number of women who were "bank tellers," "bank tellers and active feminists," and "high school teachers." The second group estimated the number of women who were "bank tellers" and "high school teachers," and the third group estimated the number of women who were "bank tellers and active feminists" and "high school teachers."

The first group provided a within-subjects test of conjunction effects, and the second and third groups provided a be-

tween-subjects test. In the within-subjects test, the majority of responses were consistent with the conjunction rule. But in the between-subjects test, responses violated the conjunction rule; the average estimate of "bank tellers and active feminists" was significantly larger than the average estimate of "bank tellers." Kahneman and Tversky argued that participants rely on the representativeness heuristic in between-subjects designs even when information is provided in frequency formats. In within-subjects designs, participants are more likely to detect set inclusion and override the heuristic.

Hertwig (1997) challenged Kahneman and Tversky's (1996) conclusion by noting that natural language terms such as "and" can be semantically ambiguous. This is crucial because in a fair test, the phrase "and" must have the same meaning in probability theory as in natural language. In probability theory, "and" refers to an intersection, but in natural language, "and" can be either a union or an intersection of events. For example, the sentence "We invited friends and colleagues to the party" implies a union, not an intersection, of friends and colleagues. See Einhorn and Hogarth (1986) and Birnbaum, Anderson, and Hynan (1990) for further discussion of the semantic ambiguity of the "and" operator.

Hertwig (1997) also pointed out that Kahneman and Tversky used different phrases when investigating probabilities and frequencies. With probabilities, Tversky and Kahneman (1983) had used the phrase "Linda is a bank teller and is active in the feminist movement," a noun-verb combination. But with frequencies, Kahneman and Tversky (1996) gave participants the phrase "bank tellers and active feminists," a noun-noun combi-

nation. Hertwig argued that the semantic ambiguity of “and” is particularly pronounced with noun-noun combinations. If some of the participants had interpreted “and” as a union rather than an intersection, Kahneman and Tversky’s (1996) between-subjects test of the conjunction rule would be invalid.

Hertwig (1997) suggested that the phrase “bank tellers who are active in the feminist movement” would make the union interpretation unlikely. He further noted that Tversky and Kahneman (1983) had used a similar phrase with probabilities. Hertwig then conducted two between-subjects tests of conjunction effects. He gave three groups of students from the University of Munich the Linda problem used by Kahneman and Tversky (1996). The first group estimated the number of women fitting the description of Linda who were “high school teachers” and “bank tellers.” The second group estimated the number of women who were “high school teachers” and “bank tellers and active feminists.” The third group estimated the number of women who were “high school teachers” and “bank tellers who are active in the feminist movement.” The first and second groups provided a between-subjects test of the word “and,” and the first and third groups provided a between-subjects test of “who are.”

Hertwig (1997) replicated Kahneman and Tversky’s (1996) between-subjects results with “and”; the average estimated number of women who were “bank tellers and active feminists” was significantly larger than the average number of women who were “bank tellers.” Hertwig also found that the average number of women who were “bank tellers who are active in the feminist movement” did not differ from the average number of women who were “bank tellers.”

Based on these results, he argued that the conjunction effects of Kahneman and Tversky (1996) may have occurred because at least some of the participants had interpreted “and” as a union. With the less ambiguous phrase “who are,” the union interpretation was less likely and conjunction effects vanished.¹ Hertwig also observed that 30% of participants who judged “bank tellers and active feminists” spontaneously asked whether the phrase referred to the intersection or the union of events. The experimenter told them to use the interpretation that seemed most appropriate. Hertwig concluded that apparent violations of the conjunction rule with frequency formats in a between-subjects design could appear or disappear, depending on the interpretation of the conjunction phrase.

At this point, our collaboration began. Kahneman conceded that “and” can be semantically ambiguous, but he did not believe that participants in Kahneman and Tversky’s (1996) study had interpreted “and” as a union. He argued that participants represented the conjunction as a prototype, as implied by the re-

presentativeness heuristic. Kahneman also thought that the phrase “bank tellers who are active in the feminist movement” is restrictive and strongly cued participants to interpret the conjunction as an intersection.

Hertwig proposed an alternative conjunction phrase, “bank tellers and are active in the feminist movement,” that he thought would also make the union interpretation unlikely. Hertwig noted that this phrase mapped Tversky and Kahneman’s (1983) original phrase, “Linda is a bank teller and is active in the feminist movement,” into a frequency format.

Kahneman accepted “and are,” but predicted that it would yield conjunction effects. He also proposed the phrase “feminist bank tellers,” which removed the conjunction phrase entirely, and predicted that it would also yield conjunction effects. We jointly designed three experiments to test these hypotheses, and Mellers collected the data. We now present methods and a brief summary of the results. Interpretations of the results are presented in the Discussion.

METHOD

Two personality sketches, one of Linda and the other of James, were used in the three experiments. The Linda story read:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Of 100 people like Linda, how many are . . . ?

The James story read:

James grew up in a Bohemian family. His father was a musician, and his mother was a painter. They lived together for 40 years and never got married. James was a very talented child with a special gift for comedy, but he turned into a rebellious troublemaker in his youth. He dropped out of college after two years and traveled to Asia to learn crafts. James is now 35 years old. Of 100 people like James, how many are . . . ?

Instructions said:

We are interested in the judgments and inferences that people make about other peoples’ professions, politics, and hobbies. In each of the following problems, we will tell you about a person. We will then ask, of 100 people like the target person, how many would fit a particular description of a job, political persuasion, or hobby? Please state your best guess when answering the following questions.

In each story, there was one likely target item and one unlikely target item. All three experiments used the target items “bank tellers” and “feminists” in the Linda story and “artists” and “Republicans” in the James story. The conjunction phrases “and” and “and are” appeared in all three experiments. A few other conjunction phrases were included in Experiments 1 and 2, as shown in Table 2. Filler items were manipulated across experiments.

1. Hertwig found that students estimated an average of 132 women out of 1,000 were “bank tellers,” 337 were “bank tellers and active feminists,” and 125 were “bank tellers who are active in the feminist movement.” Corresponding medians were 95, 250, and 23. Standard deviations were 149, 293, and 171, respectively, and the size of the sample in each group was 30.

Table 2. Average frequency estimates for all experiments

Kind of target or form of conjunction	Linda story				James story			
	Target item or conjunction phrase	Experiment			Target item or conjunction phrase	Experiment		
		1	2	3		1	2	3
Likely target	Feminists	58.1 (2.4)	47.7 (3.4)	47.9 (4.5)	Artists	41.0 (2.7)	45.1 (2.6)	47.1 (3.3)
Unlikely target	Bank tellers	24.6 (1.9)	21.4 (2.0)	14.3 (2.9)	Republicans	28.9 (2.1)	19.8 (1.8)	12.7 (2.6)
“and”	“and”	39.9 (2.0)	30.4 (2.3)	26.4 (3.9)	“and”	33.1 (1.8)	42.7 (2.4)	22.9 (3.4)
“and are”	“and are”	40.2 (2.7)	21.8 (2.1)	22.8 (2.7)	“and are”	32.0 (2.5)	20.0 (1.9)	21.4 (2.7)
[Unlikely target]	Bank tellers							
“who are”	who are				Republicans			
[likely target]	feminists	34.6 (2.3)	23.1 (2.2)	—	who are artists	26.5 (2.2)	24.0 (2.6)	—
[Likely target]	Feminists				Artists			
“who are”	who are				who are			
[unlikely target]	bank tellers	27.6 (2.2)	—	—	Republicans	29.5 (2.5)	—	—
Targets combined	Feminist							
with no conjunction	bank tellers	32.3 (2.3)	—	—	Republican artists	26.0 (2.2)	—	—

Note. Standard errors are in parentheses. Boldface indicates significant results, $p < .05$.

In Experiment 1, each participant judged the frequency of one target item and one filler item for each problem. Target items varied across groups. Filler items were “high school teachers” for Linda and “unemployed” for James. Undergraduates at Ohio State University completed the task as part of a larger take-home questionnaire, and received credit toward a course requirement for their participation. An average of 108 participants served in each group.

In Experiment 2, each participant judged the frequency of one target item for each problem; no filler items were included. Target items varied across groups. A third problem was added in that experiment (adapted from the Bill problem in Tversky & Kahneman, 1983). Ohio State University undergraduates were paid \$1 for completing the task with three problems. An average of 96 participants served in each group.

In Experiment 3, each participant judged the frequency of one target item and five filler items (shown in Table 3) for each problem. Target items varied across groups. Instructions read:

We would like you to answer these questions in two stages. First, rank the categories from most to least frequent. Write a “1” next to the most frequent category, a “2” next to the second most frequent one, etc. When you have done the ranking, please estimate how many of the 100 people belong to each category.

Undergraduates at the University of California at Berkeley volunteered to complete the questionnaire in class. An average of 40 participants served in each group.

RESULTS

Table 2 presents selected results from the three experiments. Average frequency estimates for conjunction phrases and target items are shown for Linda and James, with standard errors in parentheses. Of interest was whether the average frequency es-

timate for each conjunction was significantly greater than the average frequency estimate for the unlikely target item. Tests were done with an alpha level of 5%, and significant effects are shown in bold.

Results from the Linda problem in Experiment 1 show that conjunction effects occurred with all of the conjunctive phrases except one, “feminists who are bank tellers.” The results replicated and extended those of Kahneman and Tversky (1996). Conjunction effects occurred not only with “and,” but also with “and are” and “who are.” Furthermore, results for “and” and “and are” were no different from each other. Finally, conjunction effects occurred with “feminist bank tellers.”

In stark contrast, the James story revealed no conjunction effects whatsoever. Even the phrase “Republicans and artists” did not produce a conjunction effect. In hindsight, we believed there was less opportunity for conjunction effects to emerge in this problem because the difference between the estimates of

Table 3. Filler items and average ranks in Experiment 3

Filler item	Rank
Linda story	
Members of the League of Women Voters	2.5
Psychiatric social workers	2.8
Teachers in elementary school	3.4
Work in a bookstore and take Yoga	4.2
Insurance people	4.9
James story	
Use recreational drugs frequently	2.6
Often meditate	3.0
Write fiction and drive cabs	3.4
Work in sales and are vegetarians	4.1
Work in banking	5.1

the target items (artists vs. Republicans) was only 12.1. With Linda, the difference between the estimates of the target items (feminists vs. bank tellers) was 33.5.

We further suspected that the selection of filler items had contributed to the differences between Linda and James. With Linda, the average frequency estimate for the filler item, "high school teachers," was smaller than that of the more likely target item, "feminists" (34.8 vs. 58.1). However, with James, the average frequency estimate of the filler item, "unemployed," was larger than that of the more likely target item, "artists" (55.3 vs. 41.0). When the average estimate of the filler item was larger than that of the more likely target item, the difference between estimates of target items appeared to shrink. To avoid this problem, Hertwig proposed that Experiment 2 examine conjunction effects with no filler items.

Results for Experiment 2 show that conjunction effects occurred with "and," but vanished with "and are" and "who are." The average frequency estimate for "and" was significantly larger than that for "and are." Precisely the same pattern of results was observed for the Bill problem (adapted from Tversky & Kahneman, 1983). Hertwig's predictions for Experiment 1 were now confirmed in Experiment 2.

Kahneman suspected that participants adopted an analytic approach when judging a conjunction item in isolation. When a filler item was present, the implicit requirement to compare the conjunction item with it favored the wholistic approach that the representativeness heuristic requires. He proposed that Experiment 3 examine conjunction effects with multiple filler items and a ranking instruction prior to the estimation instruction, after Tversky and Kahneman (1982).

Results for Experiment 3 show that, with multiple filler items, conjunction effects occurred with "and" and "and are" for both Linda and James. Furthermore, the results for "and" and for "and are" did not differ.

DISCUSSION

We could easily imagine more experiments, but we had agreed that each party would be allowed one study after Experiment 1. We did not think the experiments would resolve all the issues, nor did this miracle occur. Nevertheless, we discovered more—and made the discoveries more enjoyably—than if we had engaged in replies and rejoinders. Because Hertwig and Kahneman predicted different outcomes for "and are," we focus our discussion on that phrase. Our most striking finding—which none of us had anticipated—was that "and" and "and are" produced similar results in Experiment 1 (for the Linda problem) and Experiment 3, but different results in Experiment 2.

Hertwig and Kahneman agree on some points. They agree that when filler items are present, responses to target items are implicitly comparative and more likely to produce conjunction effects. Hertwig and Chase (1998) offered a similar interpretation of target items in the presence of filler items. The interpretation is also supported by the finding that, both in the Munich

data and in our own results, conjunction effects were more likely to occur when participants gave responses that summed to 100%—clear evidence of a comparison.² In our experiments and the Munich experiment, participants were never told, either implicitly or explicitly, that their frequency estimates for the conjunction and filler items should sum to 100%. Some adopted this convention, but others did not. In Experiment 1, approximately two thirds of participants gave responses that summed to 100%, whereas in the Munich experiment, only one third gave responses that summed to 100%. Hertwig and Kahneman also agree that different psychological processes occur with comparative responses and noncomparative responses. However, they disagree about the nature of those processes. Hertwig and Kahneman now present their own perspectives.

Hertwig's View

Between-subjects designs are strange animals. They can lead to paradoxical, even absurd, inferences (e.g., Birnbaum, 1982, 1999). I agree with Varey, Mellers, and Birnbaum (1990), who argued that "one should be extremely cautious when drawing inferences from between-subjects comparisons of judgments" (p. 623). This advice is especially pertinent when one aims to investigate adherence to a rule of internal consistency, such as the conjunction rule. In our experiments, the filler items strongly affected the results obtained. One should keep in mind that filler items are superfluous in between-subjects designs in which each person sees one of the two single events or their conjunction, but never both. But when filler items are present, how do they interfere with people's judgments for the target items? One possibility is that they induce comparisons that reflect the relative support for one category (e.g., high school teachers) over the other (e.g., bank tellers). If this were the case, then a first step toward a better understanding of the dynamics of comparative processes would be to explore how the presence of specific filler items affects the estimates for the target items.

Take Experiment 3 as an example. Here, the presence of filler items appears to have differentially interfered with the estimates for the target items. Relative to Experiment 2, in which no filler items were present, the filler items in Experiment 3 dramatically reduced the frequency estimates for "bank tellers." At the same time, however, the estimates for the conjunction "bank tellers and are feminists" and for the second single event, "feminists," remained essentially unchanged (see Table 2; the same pattern is evident in the James problem). Of course, smaller estimates for "bank tellers" combined with essentially

2. Hertwig (1997) found that when responses to the target item and the filler item summed to 100%, conjunction effects occurred with "and" and "who are." When responses did not sum to 100%, both effects vanished. Results from Experiment 1 were somewhat similar. When responses to the target and filler items summed to 100%, conjunction effects occurred with "and," "who are," and "and are." But when responses did not sum to 100%, conjunction effects vanished with "and" and "who are." However, they continued to occur with "and are."

unchanged estimates for “bank tellers and are feminists” increase the likelihood of finding conjunction effects. In other words, by choosing specific filler items, it appears to be possible to make conjunction effects appear or disappear. This conclusion is also in line with the findings suggesting that specific filler items can reduce or increase the differences between the estimates for the two single-event target items (see Experiment 1).

Implicit comparative processes may also explain why the phrase “who are” yielded a conjunction effect in Experiment 1 (with the Linda problem) but not in the Munich experiment. In the Munich experiment, only a minority of estimates for the conjunction item and filler item summed to 100%, but in Experiment 1, a majority did. This is a crucial difference. Both the Munich data and the data in Experiment 1 show that conjunction effects were more likely to occur when judgments summed to 100%.

Returning to the issue on which we initially disagreed—the semantic ambiguity of natural language terms such as “and”—the ambiguity of “and” does a good job of explaining the findings in Experiment 2, in which no filler items were present and only the semantically ambiguous “and” produced conjunction effects. Nevertheless, I agree with Kahneman that semantic ambiguity appears to have played less of a role in Experiments 1 and 3. But I also believe that it is too early to conclude that semantic ambiguity plays no role in the context of implicit comparisons. It may do so in a subtler way. For instance, in the Munich experiment, “and” more often led to judgments that summed to 100% than did “who are.”

Where do we go from here? Next to semantic ambiguity of the “and” operator—an issue that, in my view, is neglected in judgment and decision-making research, which often employs text problems—the most interesting finding in our experiments is the impact of comparative processes. To shed light on their dynamics, future experiments may attempt to clarify how the content of filler item and target item interacts so that conjunction effects appear or vanish. Another interesting question is to what extent the finding that conjunction effects disappear when the conjunction is judged in isolation generalizes to representations other than frequencies. In other words, will the conjunction effect also disappear under this condition in a probability representation?

Kahneman's View

It is useful to make two familiar distinctions when discussing prediction by representativeness:

- *Extensional versus intensional* formats for the mental representation of categories: An extensional representation identifies a category by its membership; an intensional representation identifies a category by the properties of its prototypical member.
- *Intuitive versus deliberate* modes of thinking: Intuitive thinking is perception-like, rapid, effortless, and generally intensional. Deliberate thinking is reasoning-like, critical, and analytic; it is also slow, effortful, controlled, and rule-governed (Chaiken &

Trope, 1999; Stanovich & West, 2000). The processing of extensional representations appears to require deliberate thinking.

In these terms, the hypothesis of prediction by representativeness involves two separate assumptions. The first—amply supported by research in social cognition—is that prototypes are routinely evoked by mentions of social categories. The second is that judgments of probability or frequency in Linda-like problems are based on assessments of similarity between a personality description and a category prototype. The most direct evidence for the representativeness hypothesis is the fact that the rankings of outcomes by probability and by representativeness have repeatedly been shown to be identical (Kahneman & Tversky, 1973; Tversky & Kahneman, 1982). A glance at the judgments of filler items in Table 3 should convince the reader that frequency judgments follow a similar rule.

Tversky and I believed that prototype representations normally govern judgments. The baseline case is therefore that conjunction errors will occur, even in the frequency format, unless blocked by powerful extensional cues, such as the within-subjects design provides (Kahneman & Tversky, 1996; Tversky & Kahneman, 1983). In contrast, Hertwig proposed that the baseline case is that conjunction errors will not occur in the frequency format. Implicitly assuming an extensional representation, he proposed that conjunction errors arise when “and” is interpreted as a union operator, and that the results Tversky and I reported (Kahneman & Tversky, 1996) were an artifact of semantic ambiguity.

The semantic-ambiguity hypothesis was tested and conclusively rejected in Experiments 1 and 3, in which the effects of the unambiguous “and are” were undistinguishable from the effects of “and.” However, the results of the instructive experiment that Hertwig proposed came as a complete surprise to me: Conjunction effects completely disappeared with “and are” in Experiment 2, whereas the effects for “and” were stable across all experiments.

There is no reason to believe that the union interpretation of “and” played a role in Experiment 2, because it did not in the other experiments. The odd result that requires explanation is why the presence or absence of fillers affected judgments of “and are.” I infer from this finding that (a) the implicit instruction to compare the target item to the filler item favored a unified prototype representation of the conjunction category in Experiments 1 and 3, for “and are” as well as for “and,” and (b) the exclusive focusing of attention on a single item in Experiment 2 favored a deliberate analytic approach, in which respondents were able to detect the restrictive implication of “and are.”

This admittedly post hoc interpretation is readily testable. For example, it entails that conjunction errors with “and are” should be restored by replicating Experiment 2 under mild cognitive load, or after a positive-mood manipulation (Schwarz & Clore, 1996). Conversely, conjunction errors should be reduced—perhaps even with “and”—by replicating Experiment 1 (with filler items) under conditions that focus participants’ at-

tention on the language of the items, by requiring respondents to read them aloud, or by using an earlier task to prime an analytic approach. There are other ways of extending the empirical study of the issues we raised. For example, the initial issue on which we disagreed—whether respondents interpret “and” as a union operator—can be tested directly by asking them to indicate their chosen interpretation on a Venn diagram. And an issue on which our views are converging—the role of comparative strategies as moderators of conjunction effects—also invites further empirical exploration.

CONCLUSION

Our joint efforts demonstrate the benefits of adversarial collaboration as a method for conducting scientific controversy. The major benefit is that both parties are likely to recognize limitations of their claims. In our case, Hertwig acknowledges that the semantic ambiguity of “and” is only one factor that contributes to conjunction effects in the frequency format. Kahneman acknowledges that the experiments have identified conditions under which the representativeness heuristic is overriden, and that there may be more to be discovered.

Collaborating with an adversary is not easy. In Table 1, we offer suggestions that would have made our experience better. Despite our mishaps, we hope the approach catches on. In an ideal world, scholars would feel obliged to accept an offer of adversarial collaboration. Editors would require adversaries to collaborate prior to, or instead of, writing independent exchanges. Scientific meetings would allot time for scholars engaged in adversarial collaboration to present their joint findings. In short, adversarial collaboration would become the norm, not the exception.

REFERENCES

- Adler, J. (1984). Abstraction is uncooperative. *Journal for the Theory of Social Behaviour*, 14, 165–181.
- Birnbaum, M.H. (1982). Controversies in psychological measurement. In B. Wegner (Ed.), *Social attitudes and psychophysical measurement* (pp. 401–485). Hillsdale, NJ: Erlbaum.
- Birnbaum, M.H. (1999). How to show that $9 > 221$: Collect judgments in a between-subjects design. *Psychological Methods*, 4, 243–249.
- Birnbaum, M.H., Anderson, C.J., & Hynan, L.G. (1990). Theories of bias in probability judgment. In J.-P. Caverni, J.-M. Fabre, & M. Gonzalez (Eds.), *Cognitive biases* (pp. 477–498). Amsterdam: North Holland.
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York: Guilford Press.
- Dulany, D.E., & Hilton, D.J. (1991). Conversational implicature, conscious representation and the conjunction fallacy. *Social Cognition*, 9, 85–110.
- Einhorn, J.H., & Hogarth, R.M. (1986). Judging probable cause. *Psychological Bulletin*, 99, 3–19.
- Fiedler, K. (1988). The dependence of the conjunction fallacy on subtle linguistic factors. *Psychological Research*, 50, 123–129.
- Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases.” *European Review of Social Psychology*, 2, 83–115.
- Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky (1996). *Psychological Review*, 103, 592–596.
- Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, 102, 684–704.
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98, 506–528.
- Gilovich, T., Medvec, V.H., & Kahneman, D. (1998). Varieties of regret: A debate and partial resolution. *Psychological Review*, 105, 602–605.
- Hertwig, R. (1997). *Judgment under uncertainty: Beyond probabilities*. Unpublished manuscript, Max Planck Institute for Human Development, Berlin, Germany.
- Hertwig, R., & Chase, V.M. (1998). Many reasons or just one: How response mode affects reasoning in the conjunction problem. *Thinking and Reasoning*, 4, 319–352.
- Hertwig, R., & Gigerenzer, G. (1999). The ‘conjunction fallacy’ revisited: How intelligent inferences look like reasoning errors. *Journal of Behavioral Decision Making*, 12, 275–306.
- Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.
- Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer’s critique. *Psychological Review*, 103, 582–591.
- Politzer, G., & Noveck, I.A. (1991). Are conjunction rule violations the result of conversational rule violations? *Journal of Psycholinguistic Research*, 20, 83–103.
- Schwarz, N., & Clore, G.L. (1996). Feelings and phenomenal experiences. In T.E. Higgins & A.W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 433–465). New York: Guilford.
- Stanovich, K.E., & West, R.F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23, 645–726.
- Tversky, A., & Kahneman, D. (1982). Judgment of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–100). New York: Cambridge University Press.
- Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.
- Varey, C., Mellers, B.A., & Birnbaum, M.H. (1990). Judgments of proportions. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 613–625.

(RECEIVED 7/26/00; ACCEPTED 8/25/00)