Narrative Impressions of Literature: The Availability Bias and the Corrective Properties of Meta-Analytic Approaches

Brad J. Bushman
Gary L. Wells
Iowa State University

*Participants (*N = *280) reviewed 20 fictional research summaries of studies examining the relation between similarity and attraction. Although there were some inconsistencies across the fictional studies, there was a positive relation overall (*d = *0.2). The authors manipulated the salience of the titles and the serial order in which the studies were presented without changing the results of the studies themselves. Participants recalled the salient titles better than the nonsalient titles. Participants who were given very brief training in meta-analytic techniques gave estimates of the similarity-attraction relation that were close to the actual magnitudes. Participants who were not given such training (narrative reviewers) were influenced by the salience manipulation and gave estimates that were biased toward the studies that had salient titles. Although the salience manipulation influenced participants in the meta-analytic and narrative groups equally, memory mediated the effects of salience on estimates of effect magnitude only for the narrative review participants.*

It is somewhat ironic that the traditional review of scientific data has typically been conducted in an unscientific fashion. In the traditional narrative review, the reviewer uses "mental algebra" to combine the findings from a collection of studies and describes the results verbally. Statisticians were the first to advocate alternative methods for combining research findings. These methods were labeled meta-analysis by Gene Glass (1976):

> Meta-analysis refers to the analysis of analyses . . . the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding literature. (p. 3)

The quantification of research evidence is the key factor that distinguishes a meta-analytic review from a narrative review (Olkin, 1990). In the meta-analytic review, the meta-analyst uses statistical procedures to integrate the findings from a collection of studies and describes the results using numerical effect-size estimates.

Narrative reviews suffer from at least two weaknesses. First, narrative reviews can be more susceptible to the subjective judgments, preferences, and biases of a particular reviewer's perspective than meta-analytic reviews. Second, narrative reviews ignore the magnitude of the treatment effect. In a narrative review, the reviewer frequently uses *p* values to draw conclusions by counting the number of studies that found significant treatment effects. These two weaknesses of narrative reviews can cause their conclusions to be inconsistent with the data.

In a study by Cooper and Rosenthal (1980), for example, faculty members and upper level graduate students in psychology were randomly assigned to use narrative or meta-analytic procedures to review seven studies on sex differences in persistence. None of the reviewers were familiar with meta-analytic procedures. Participants in the meta-analysis group were instructed on how to combine the *p* values from the studies. Participants in the narrative group were instructed to "employ whatever criteria you would use if this exercise were being undertaken for a class term paper or a manuscript for publica-

Authors' Note: We would like to thank Craig Anderson and Harris Cooper for their helpful comments on an earlier version of this article. We also would like to thank Kaila Boothroy and Laurie Parcher for serving as experimenters and Janelle Johnson and Denise Allumbaugh for serving as raters. Correspondence concerning this article should be addressed to Brad J. Bushman, Department of Psychology, Iowa State University, Ames, IA 50011-3180; e-mail: bushman@iastate.edu.

PSPB, Vol. 27 No. 9, September 2001 1123-1130

tion." Participants were asked whether the evidence supported the conclusion that women were more persistent on tasks than were men. Five possible responses were provided (*definitely yes*, *probably yes*, *impossible to say*, *probably no*, or *definitely no*). The results showed that 68% of the meta-analytic reviewers were at least considering rejecting the null hypothesis, compared with only 27% of the narrative reviewers. (The null hypothesis should have been rejected because the combined p value was less than .05.) Participants also were asked to estimate the magnitude of sex differences in persistence. Six possible responses were provided (*very large*, *large*, *moderate*, *small*, *very small*, or *none at all*). The results showed that 58% of the meta-analytic reviewers estimated at least a small sex difference in persistence, compared with only 27% of the narrative reviewers. (The effect was about equal to Cohen's [1988] conventional value for a "small" effect.) Thus, participants in the narrative group underestimated the presence and the strength of sex differences in persistence.

Similar results have been reported elsewhere. For example, an article in *Science* (Mann, 1994) compared the conclusions drawn from meta-analytic versus traditional literature reviews in five subject areas: (a) psychotherapy, (b) delinquency prevention, (c) school funding, (d) job training, and (e) reducing anxiety in surgical patients. The comparison revealed that narrative reviews underestimated the presence and the strength of treatment effects for each subject area.

Psychological Factors Influencing Narrative Reviews

Mistrust of conclusions reached through narrative reviews, relative to meta-analytic reviews, is likely to be well founded on a number of counts. One factor that might contribute to bias from narrative reviews is the reviewer's preconceived notions about relations among the variables being studied. Preconceptions are likely to persevere in the face of contradictory data when the new data are not clear-cut (Anderson & Sechler, 1986). Likewise, preconceptions tend to be particularly powerful in influencing human judgment when the information is mixed and subject to multiple conclusions (Hilton & von Hippel, 1990). Interestingly, social cognition research shows that rather than ignoring information inconsistent with preconceptions, people sometimes pay extra attention to such information (Hilton, Klein, & von Hippel, 1991) but then go on to discount the information using any number of psychological mechanisms, such as the confirmation bias (e.g., Lord, Ross, & Lepper, 1979). The confirmation bias is the tendency to search for information that confirms one's prior beliefs. In the case of narrative reviews, for instance, we suspect that there are natural propensities to search rigorously for methodological flaws in the studies that do not fit the reviewer's prior beliefs and engage in a much less vigorous search for flaws in studies that match the reviewer's prior beliefs. When people have prior theories about the relations among variables, several processes (e.g., behavioral confirmation, biased attribution and recall, and biased assimilation) can produce systematic data distortions, such as illusory correlations (Anderson & Lindsay, 1998).

Our focus is on another possible problem with narrative reviews; namely, that judgments of the likelihood or frequency of a finding might be influenced by the ease with which relevant instances come to mind. The size and complexity of most literatures for which one directs a narrative review undoubtedly exceed the reviewers' ability to perform accurate mental arithmetic; hence, the impressions or conclusions reached from reading a literature are likely to be influenced by heuristics or mental shortcuts. A pervasive heuristic in human judgment is the availability heuristic, in which people judge the frequency or prevalence of some event by the ease with which relevant instances come to mind (Tversky & Kahneman, 1973). Judging the frequency or prevalence of an event using the availability heuristic can lead to biased judgments because the ease with which relevant instances come to mind can be influenced not only by the actual frequencies of the event but also by a host of other variables that affect its mental availability (e.g., vividness, recency, familiarity). This type of bias does not arise from preconceptions but rather from properties of the events that make them more or less easily come to mind at the time of the judgment. We were interested in testing this idea as it relates to narrative versus meta-analytic reviews.

We hypothesized that a narrative-review approach to integrating a literature would show availability bias effects, whereas a meta-analytic approach to integrating a literature would be immune to such effects. We decided to test this hypothesis by having participant-reviewers read a set of fictitious studies. We manipulated two variables that we thought would affect the availability of one particular finding over another. The two variables that we chose to manipulate were the salience of the titles of the studies and the order in which the studies were sequentially encountered. We assumed that some titles (e.g., "Birds of a Feather Flock Together") would be more memorable than other titles (e.g., "Research Examines Similarity as Source of Liking"), a manipulation that we will refer to as title salience. We held constant the actual research results reported under those titles. This is analogous in key theoretic respects to classic demonstrations of the availability bias, such as the fame-and-gender effect (see McKelvie, 1995). We

expected that studies with salient titles would be better recalled and, hence, the findings reported under those titles would be overrepresented in the participant-reviewers' conclusions about the relations among the variables being studied.

We also manipulated the order in which the studies were encountered by the participant-reviewers, hoping to take advantage of serial position effects, which typically show that items early in a sequence and late in a sequence are better recalled than are items toward the middle of a sequence (Murdock, 1962). Hence, given a mixture of results across studies (some showing positive effects, others not), placing most of the positive effects at the beginning and end of the sequence of studies could lead to different impressions of the results overall than placing those studies in the middle of the sequence.

METHOD

Participants

Participants were 280 undergraduate students (140 men and 140 women) who received extra course credit in exchange for their voluntary participation. None of the participants were familiar with meta-analytic procedures.

Materials

We created a fictional set of 20 research summaries for the participants to read. Each of the studies reported the results of research on the question of whether interpersonal attraction is related to similarity. A sample research summary is given below.

> In a recent study, 386 high school students were asked to list the people they liked most and least in their class. The students also listed their hobbies and interests. Students had many more hobbies and interests in common with the people they liked most (3.2, on average) than with the people they liked least (0.8, on average), $t(384) = 10.12$, $p < .05$. These findings suggest that similarity is an important source of liking.

To avoid confusing participant-reviewers, we did not report exact p values. For studies with nonsignificant results, we used $p > .05$. In line with the actual pattern of findings in social psychology, the results across studies were mixed but the overall relation across the studies was one in which similarity was positively related to attraction (Berscheid, 1985).

The standardized mean difference, denoted by d, was used to quantify the magnitude of effects in the research summaries; it gives the number of standard deviation units between two groups. The standardized mean difference for the ith study was

$$d = \frac{M_{\text{SIMILAR}} - M_{\text{DISSIMILAR}}}{SD}$$

where M_{SIMILAR} and $M_{\text{DISSIMILAR}}$ are the respective sample means for the similar and dissimilar groups and SD is the pooled sample standard deviation of the two groups (Cooper, 1989, pp. 101-102). An effect was defined as "positive" if the level of attraction for the similar group was significantly greater than the level of attraction for the dissimilar group. An effect was defined as "negative" if the level of attraction for the similar group was significantly less than the level of attraction for the dissimilar group.

The sample effect sizes for the 20 studies were generated from a normal distribution with mean 0.20 and variance 0.20 using the Statistical Analysis System (SAS) RANNOR function. We purposely used a large variance because we wanted some studies to have moderate-sized negative effects. Sample effect sizes ranged from 1.03 to –0.62. The weighted average of 20 effects was 0.21, with a 95% confidence interval ranging from 0.16 to 0.25. Because the confidence interval does not include the value zero, the average effect is greater than zero. The magnitude of the average effect is similar to Cohen's (1988) conventional value for a "small" effect (i.e., $d = 0.20$). This effect is similar in magnitude to many effects observed in social and personality psychology (e.g., Bond & Titus, 1983; Eagly & Carli, 1981; Fejfar & Hoyle, 2000; Gerrard, Gibbons, & Bushman, 1996; Gordon, 1996; Lytton & Romney, 1991; Oliver & Hyde, 1993; Rawsthorne & Elliot, 1999; Sheeran, Abraham, & Orbell, 1999; Swim & Sanna, 1996).

The total sample sizes for the 20 studies were generated from a normal distribution with mean 400 and variance 400 using the SAS RANNOR function. The total sample sizes were constrained to be even integers so the sample sizes for the two groups would be equal in each study. Total sample sizes ranged from 428 to 378 ($M = 400.1$, $SD = 14.2$).

The independent t test for the ith study was computed as

$$t = \frac{d\sqrt{N-2}}{2},$$

where N is the total sample size for the ith study (Friedman, 1968). Of the 20 studies, 9 had positive results, 3 had negative results, and 8 had null results.

Procedure

Individuals participated in small groups but worked independently on the experimental tasks. Participants were told that the researchers were studying how people combine information from research studies. After informed consent was obtained, participants received a

packet containing 20 summaries of (fictitious) studies of the relation between similarity and attraction.

Four variations of the research summary packet were created by manipulating two factors. One factor was whether salient titles were attached to positive or negative results. Examples of salient titles for positive results include "Birds of a Feather Flock Together," "From the Same Mold," and "Peas in a Pod." Examples of salient titles for negative results include "Opposites Attract," "Different as Night and Day," and "Nothing in Common." Examples of nonsalient titles include "Research Examines Similarity as Source of Liking," "Social Psychologists Study Matchmaking," and "Research Asks Who Likes Whom." Because we wanted the title salience manipulation to affect perceptions of the direction of the effect, we did not manipulate the salience of titles for the null results. Null results always had nonsalient titles. The other manipulated factor was whether studies with positive results appeared in the beginning and end of the packet or in the middle of the packet. Participants were randomly assigned to receive one of the four types of packets.

After participants read the research summaries and turned in their packets, they were given a surprise free recall test in which they recalled as many of the 20 titles as possible. These data were used as a manipulation check to determine whether salient titles were more easily recalled than nonsalient titles and whether titles appearing in the beginning and end of the packet were more easily recalled than titles appearing in the middle of the packet. These data also allowed us to test whether narrative and meta-analysis groups differed in their recall of titles. Because participants were randomly assigned to meta-analytic and narrative review groups, we did not expect their recall to differ.

Participants were then randomly assigned to narrative or meta-analytic review conditions. The experimenter returned the packets and asked participants to carefully review the evidence in the 20 research summaries. In the narrative group, participants were told, "For your review, employ whatever criteria you would use if this exercise were being undertaken for a class term paper or a manuscript for publication." Participants in the meta-analysis group received a calculator and brief instruction on how to compute and combine effect-size estimates. Participants in the meta-analysis group computed effect-size estimates from t tests using Friedman's (1968) formula and then simply averaged effect-size estimates. To simplify calculations, we decided not to have participants in the meta-analysis group weight each effect-size estimate by the inverse of its variance or compute a confidence interval for the combined effect-size estimate. Because the 20 fictitious studies had similar sample sizes, the weighted and unweighted combined effect-size estimates were similar (0.21 and 0.22, respectively). Participants in the meta-analysis group were told to interpret effect-size estimates using Cohen's (1988) conventional values for "small" ($d = 0.20$), "medium" ($d = 0.50$), and "large" ($d = 0.80$) effects. Participants in the meta-analysis group also were asked to show their work so that we could check whether they computed and combined effect-size estimates correctly.

There were two measures that combined to form the main dependent variables. The first measure was, "The evidence presented, in general, supports which conclusion?" Three options were given: (a) similarity increases attraction (coded +1), (b) similarity decreases attraction (coded –1), and (c) similarity has no effect on attraction (coded 0). The second measure was, "Estimate the magnitude of the effect of similarity on attraction." Six options were given: (a) *none at all* (coded 0), (b) *very small* (coded 1), (c) *small* (coded 2), (d) *moderate* (coded 3), (e) *large* (coded 4), and (f) *very large* (coded 5). After responding to each of these two measures, participants were fully debriefed and dismissed.

RESULTS

Sex Differences

Participant sex did not influence any of the results, alone or interacting with other variables. Thus, the data from men and women were combined for subsequent analyses.

Reliability of Coding

Recall Measures

Two independent raters, who were blind to conditions and experimental hypotheses, coded the titles recalled. Raters coded whether the title was salient or nonsalient; whether the title appeared in the beginning/end of the packet or in the middle of the packet; and whether the title was attached to a study reporting a positive, negative, or null result. Agreement among raters was assessed using the intraclass correlation coefficient (Shrout & Fleiss, 1979). The coefficients ranged from .88 to .93. Because the degree of agreement between raters was high, scores from the two raters were averaged for subsequent analyses.

Manipulation Checks

As expected, participants recalled more salient titles than nonsalient titles, $t(239) = 4.58$, $p < .0001$. The magnitude of this effect was $d = 0.30$, with a 95% confidence interval ranging from 0.17 to 0.43. The order of titles also influenced recall, although the effect was not quite significant, $t(239) = 1.97$, $p = .05$, $d = 0.13$ (0.00, 0.26).

We also checked whether participants in the meta-analysis group calculated the correct average effect-size

estimate. Of the 120 participants in the meta-analysis group, 105 showed their work (88%). Of those who showed their work, 96 (91%) performed the calculations correctly.

Recall Data

As expected, there were no differences between the narrative and meta-analysis literature review groups in recall of salient titles, nonsalient titles, first and last titles, or middle titles, *t*s(238) = 0.84, 0.22, 0.54, and 0.11, respectively, *p*s > .05. Thus, the two groups did not differ in their ability to recall the titles.

Participants' Estimates of the Similarity-Attraction Relation

Participants' estimates of the similarity-attraction relation were analyzed using a three-way analysis of variance with factors of title (salient titles attached to positive or to negative results), order (positive results appeared in the beginning and end of the packet or in the middle of the packet), and literature review procedure (narrative or meta-analysis). The two dependent measures (direction of relation and magnitude of relation) were multiplied together to form an index of the relation between similarity and attraction. This produced an 11-point measure that ranged from –5 (*very large negative effect*) to +5 (*very large positive effect*). We call this new dependent variable "estimated effect magnitude."

Analysis of variance revealed a significant interaction between title salience and literature review group, $F(1, 232) = 4.80$, $p < .03$ (see Figure 1). In the narrative literature review group, the article titles had a significant effect on participants' judgments of the relation between similarity and attraction, $t(232) = 2.73$, $p < .007$, $d = 0.50$ (0.14, 0.86). The relation between similarity and attraction was judged to be stronger if studies with positive results had salient titles than if studies with negative results had salient titles. Title saliency did not influence ratings in the meta-analysis group, $t(232) = -0.37$, $p = .71$, $d = -0.07$ (–0.43, 0.29). There were no other significant results.

Memory as a Mediator

Our hypothesis was that the manipulated variables would influence recall of the studies, which, in turn, would affect overall impressions of the relation between the similarity and attraction variables in the narrative (but not the meta-analytic) conditions. However, the serial order manipulation did not affect recall. Accordingly, our hypothesis was tested with the salience manipulation. Our hypothesized model is a moderated-mediation model in which type of review (narrative vs. meta-analytic) is the moderator and memory is the mediator.

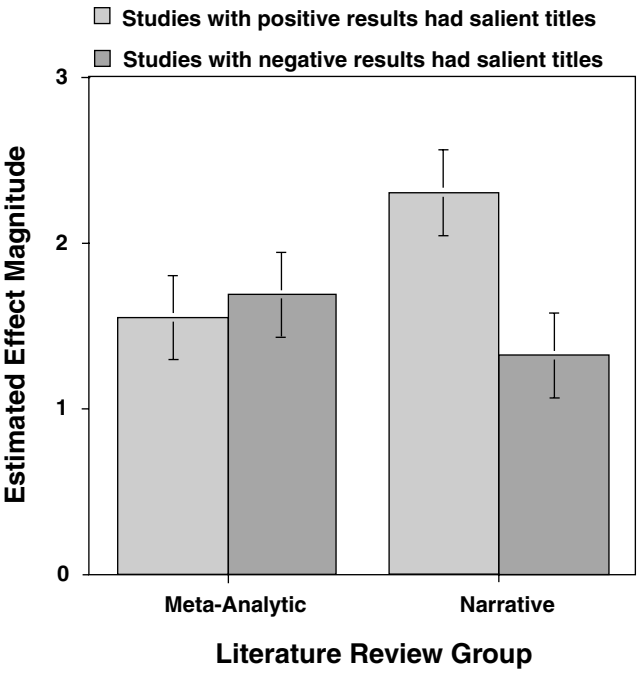

Figure 1 Effect magnitude estimates for meta-analytic and narrative literature groups as a function of whether positive or negative results had salient titles.

NOTE: Estimated effect magnitude could range from –5 (*a very large negative effect*) to +5 (*a very large positive effect*). Capped vertical bars denote 1 standard error.

To test this model, a multiple group analysis was conducted using the LISREL 8 computer program (Jöreskog & Sörbom, 1993). Table 1 gives the variance-covariance matrices for the analyses. For participants in the narrative review group, a causal path was specified from Title Salience to Memory and another causal path was specified from Memory to Estimated Effect Magnitude. For participants in the meta-analysis group, a causal path was specified from Title Salience to Memory, and the causal path from Memory to Estimated Effect Magnitude was fixed at zero. The hypothesized model provided an excellent fit to the data, $\chi^2(3, N = 240) = 2.79$, $p = .43$, Goodness of Fit Index = 1.00, Comparative Fit Index = 1.00, root mean square error of approximation = 0.043. All hypothesized pathways were significant.

A second model was tested in which all paths were freely estimated. This model also provided an excellent fit to the data, $\chi^2(2, N = 240) = 2.65$, $p = .27$ (see Figure 2). However, the fit did not improve significantly between the two models, $\chi^2(1, N = 240) = 0.14$, $p = .71$. The causal pathway between Memory and Estimated Effect Magnitude was nonsignificant in the meta-analysis group, $z = -0.37$, $p = .71$.

A third model was tested that also included a direct path from Title Salience to Estimated Effect Magnitude for both literature review groups. Once again, inclusion

TABLE 1: Data for LISREL Analysis

Measure	*1*	*2*	*3*	M	SD
Narrative review group					
1. Title salience	0.25	0.37	0.25	0.50	0.50
2. Memory	0.46*	2.52	0.72	1.57	1.59
3. Estimated effect magnitude	0.22*	.20*	4.98	1.81	2.04
Meta-analysis review group					
1. Title salience	0.25	0.38	–0.03	0.50	0.50
2. Memory	0.44*	0.25	0.25	1.59	1.68
3. Estimated effect magnitude	–0.04	–.03	2.91	1.62	1.71

NOTE: Variances are on the diagonal; covariances are above the diagonal. Correlations are reported below the diagonal for descriptive purposes. $n = 120$ in each condition.
*$p < .05$.

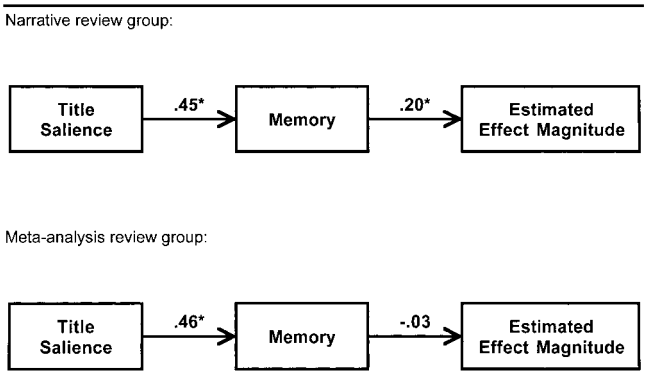

Figure 2 Memory as a mediator between title salience and estimated effect magnitude in narrative and meta-analysis review groups.
*$p < .05$.

of this parameter did not lead to a significant improvement in the fit of the model to the data, $\chi^2(1, N = 240) = 0.04$, $p = .84$. The direct causal path from Title Salience to Estimated Effect Magnitude was nonsignificant for both the narrative and meta-analytic review groups, $z = 1.61$, $p = .11$; $z = –0.43$, $p = .67$, respectively.

DISCUSSION

We have shown that narrative review approaches to making conclusions about a literature can be biased systematically through variables other than the preconceived notions of the reviewers. Although our participants might have had preconceived notions about the similarity-attraction relation, random assignment assured that the effects we observed were not due to these preconceived notions. Instead, superficial properties of the studies themselves (title salience) influenced narrative reviewers' impressions of the literature. Using salient titles for the positive results led to overestimates of the actual relation and using salient titles for the negative results led to underestimates of the actual relation. Our test of mediation for the narrative reviewers showed a path through memory that seems to explain nearly all of the effect of title salience on effect magnitude estimates. The effect of title salience on memory was equally strong for the meta-analysis participants, but their estimates of effect magnitude were unaffected by the salience manipulation and no path through the memory variable was observed.

One possible concern about the generalizability of our results is that we had our participants recall the titles before conducting their reviews. Perhaps this made the narrative review participants more biased toward the salient titles than would normally be the case. However, we think reviewers usually engage in spontaneous recall prior to writing up their conclusions; therefore, we believe that this recall task simulated a natural process at some level. Furthermore, this recall task was used for participants in both the narrative and meta-analytic conditions, which makes it especially impressive that the meta-analytic condition participants were uninfluenced by the salience manipulation.

It is unclear why manipulating the serial position of the positive and negative result studies did not affect recall. However, we note that the nature of the task for our participants was not characteristic of the traditional type of task that has shown serial position effects in free recall. Serial position effects are typically demonstrated with materials in which the items are independent and the participants are not trying to integrate across the items and there is no pattern across the items. Unlike a list of unrelated words or names, for instance, we suspect that participants were trying to integrate the serially encountered studies. Furthermore, by placing studies toward the middle that were inconsistent with those at the beginning, we might have inadvertently canceled any serial position effect by the extra attention participants might have given to the middle studies.

The relative ease with which we were able to "train" participants to base their conclusions on the meta-analytic solution is consistent with a broad body of recent research showing that even brief statistical training can improve reasoning (Nisbett, 1993). Such effects rely critically on the participant having the proper statistical rule for producing a solution, calculating that solution, and then using the statistical solution in reaching a conclusion. Our instructions gave participants the proper statistical rule and, having calculated the meta-analytic answer (91% calculated the correct answer), they used that answer in making their conclusions.

We used the title salience manipulation as a proxy for the more general idea that nonmathematical impres-

sions of a literature can be influenced by cognitive heuristics, such as availability. We do not necessarily intend to suggest that psychological scientists' narrative reviews of complex literatures are influenced by the salience of the titles of the articles encompassed by that review. Title salience might or might not lead a professional to more heavily represent one study than some other study in developing an impression of the direction or magnitude of relations among variables. On the other hand, surely it is the case that not all studies equally come to mind when reflecting on a literature, even for the professional. Possible influences on the availability of studies include the fame of the author(s), whether one knows or does not know the authors, the quality of the journal, the fluency of the text, the number of times the study was read, and so on.

Heuristics can sometimes yield very accurate solutions, solutions that are nearly identical to those derived through explicit mathematical formulas. The use of the availability heuristic in estimating frequencies, for instance, can yield very accurate solutions if salience and frequency are correlated (e.g., when the salient events are more frequent than the nonsalient events). We see no particular reason, however, to think that the salience of a study should be positively correlated with the direction or magnitude of an effect across a body of literature. In fact, the correlation between salience and direction of effect might even be negative under some circumstances. For instance, we think it is possible that studies that are exceptions to the general finding in literature will be disproportionately represented in the mental algebra of a reviewer because the exceptions are more salient than the general rule.

If our speculation is correct, exceptions should have a greater impact in narrative reviews than they do in meta-analytic reviews. Our speculation is consistent with research showing that narrative reviews seem to underestimate the presence and strength of treatment effects in psychotherapy, delinquency prevention, school funding, job training, and anxiety reduction in surgical patients (Mann, 1994). We know of no research showing that narrative reviews overestimate effect sizes compared to meta-analytic reviews. This is somewhat surprising if we consider the primary bias in narrative reviews to be that of using preconceived notions and then dismissing the inconsistencies. However, if we assume that exceptions to the general rule are more salient and hence more available in memory, then the tendency for narrative reviews to underestimate effect sizes is precisely what would be expected.

The American Psychological Association's (APA) Task Force on Statistical Inference recently issued guidelines for revising the *APA Publication Manual* (Wilkinson and the Task Force on Statistical Inference, 1999). Among the recommendations is "Always present effect sizes for primary outcomes" (p. 599). Although speculative, we think that the effect of title salience might have been muted or even eliminated if our research summaries had included effect size statistics rather than just means and standard deviations. If so, this is additional ammunition for incorporating the effect size recommendation from the Task Force on Statistical Inference into the next edition of the *APA Publication Manual*.

Psychological science has a long and distinguished tradition of studying cognitive biases and documenting the deficiencies of informal impressions by both lay people and professionals. Hence, psychologists should understand, even better than do other scientists, the problems with impressionistic, narrative reviews of the literature. For this reason, psychology should be among the most outspoken advocates of meta-analysis in all the sciences.

REFERENCES

Anderson, C. A., & Lindsay, J. J. (1998). The development, perseverance, and change of naïve theories. *Social Cognition, 16*, 8-30.

Anderson, C. A., & Sechler, E. S. (1986). Effects of explanation and counter-explanation on the development and use of social theories. *Journal of Personality and Social Psychology, 50*, 24-34.

Berscheid, E. (1985). Interpersonal attraction. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 2, 3rd ed.). New York: Random House.

Bond, C. F., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin, 94*, 265-292.

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

Cooper, H. M. (1989). *Integrating research: A guide for literature reviews* (2nd ed.). Newbury Park, CA: Sage.

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. *Psychological Bulletin, 87*, 442-449.

Eagly, A. H., & Carli, L. L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influence studies. *Psychological Bulletin, 90*, 1-20.

Fejfar, M. C., & Hoyle, R. H. (2000). Effect of private self-awareness on negative affect and self-referent attribution: A quantitative review. *Personality and Social Psychology Review, 4*, 132-142.

Friedman, H. (1968). Magnitude of an experimental effect and a table for its rapid estimation. *Psychological Bulletin, 70*, 245-251.

Gerrard, M., Gibbons, F. X., & Bushman, B. J. (1996). The relation between perceived vulnerability to HIV and precautionary sexual behavior. *Psychological Bulletin, 119*, 390-409.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3-8.

Gordon, R. A. (1996). Impact of ingratiation on judgments and evaluations: A meta-analytic investigation. *Journal of Personality and Social Psychology, 71*, 54-70.

Hilton, J. L., Klein, J. G., & von Hippel, W. (1991). Attention allocation and impression formation. *Personality and Social Psychology Bulletin, 17*, 548-559.

Hilton, J. L., & von Hippel, W. (1990). The role of consistency in the study of stereotype-relevant behaviors. *Personality and Social Psychology Bulletin, 16*, 430-448.

Jöreskog, K. G., & Sörbom, D. (1993). *Windows LISREL 8.12*. Chicago: Scientific Software.

Lord, C. G., Ross, L., & Lepper, M. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subse-

quently considered evidence. *Journal of Personality and Social Psychology, 47*, 1231-1243.
Lytton, H., & Romney, D. M. (1991). Parents' differential socialization of boys and girls: A meta-analysis. *Psychological Bulletin, 109*, 267-296.
Mann, C. (1994). Can meta-analysis make policy. *Science, 266*, 960-962.
McKelvie, S. J. (1995). Bias in the estimated frequency of names. *Perceptual and Motor Skills, 81*, 1331-1338.
Murdock, B. B., Jr. (1962). The serial position effect for free recall. *Journal of Experimental Psychology, 64*, 482-488.
Nisbett, R. E. (Ed.). (1993). *Rules for reasoning*. Hillsdale, NJ: Lawrence Erlbaum.
Oliver, M. B., & Hyde, J. S. (1993). Gender differences in sexuality: A meta-analysis. *Psychological Bulletin, 114*, 29-51.
Olkin, I. (1990). History and goals. In K. W. Wachter & M. L. Straf (Eds.), *The future of meta-analysis* (pp. 3-26). New York: Russell Sage.
Rawsthorne, L. J., & Elliot, A. J. (1999). Achievement goals and intrinsic motivation: A meta-analytic review. *Personality and Social Psychology Review, 3*, 326-344.
Sheeran, P., Abraham, C., & Orbell, S. (1999). Psychosocial correlates of heterosexual condom use: A meta-analysis. *Psychological Bulletin, 125*, 90-132.
Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428.
Swim, J. K., & Sanna, L. J. (1996). He's skilled, she's lucky: A meta-analysis of observers' attributions for women's and men's successes and failures. *Personality and Social Psychology Bulletin, 22*, 507-519.
Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*, 207-302.
Wilkinson, L., & the Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 54*, 594-604.

Received December 3, 1999
Revision accepted August 12, 2000