Frequency and the Conference of Referential Validity

LYNN HASHER AND DAVID GOLDSTEIN

Temple University

AND

THOMAS TOPPINO

Villanova University

Subjects rated how certain they were that each of 60 statements was true or false. The statements were sampled from areas of knowledge including politics, sports, and the arts, and were plausible but unlikely to be specifically known by most college students. Subjects gave ratings on three successive occasions at 2-week intervals. Embedded in the list were a critical set of statements that were either repeated across the sessions or were not repeated. For both true and false statements, there was a significant increase in the validity judgments for the repeated statements and no change in the validity judgments for the non-repeated statements. Frequency of occurrence is apparently a criterion used to establish the referential validity of plausible statements.

In the standard memory task, the learner is presented information which he is instructed to read or remember. The information may range from simple units such as nonsense syllables, numbers, and words, to complex units such as phrases and prose passages. Memory for that information may be tested using any of a wide range of procedures. From the subject's perspective these input events have what we call episodic validity. That is, the information units occur in a situation surrounded by scientific dignity, requiring responsible behavior from the subject (Orne, 1973), and also in a situation in which the subject has volunteered his services, further contributing to the subject's compliance with the demands and expectations of the experimenter (Rosenthal & Rosnow, 1975). The information units themselves, however, have little validity apart from that conferred on them by the experimental context. For example, the statement "the rock which rolled down the mountain crushed the tiny hut at the edge of the woods" in the context of a memory experiment (Bransford & Franks, 1971) has validity in the sense that it occurs in that experimental episode. The subject need not believe, nor is it even likely to occur to him, that the statement refers to a real world event in order for him to comprehend and respond to that statement.

We were curious about the kinds of processing subjects do with information units that have potential reference to the real world, that is, with items that have what we will call referential validity. Take as an example, the statement that "The total population of Greenland is about 50,000." This sentence, unlike the previous example, is plausible; that is, it is potentially verifiable in that there is a specific referent. However, it is also a statement about whose referential validity most of us would have some uncertainty. We can of course judge that such a statement might be true, presumably by using general information from semantic memory. Nevertheless, it seemed to us that people are willing to make judgments about the truth or falsity of such plausible statements in the absence of certain

Send reprint requests to Lynn Hasher, Department of Psychology, Weiss Hall, Temple University, Philadelphia, PA 19122.

knowledge. What, then, could be the basis for such judgments? This is a question not about the contents of semantic memory, nor about the relations among those contents, but rather about the fundamental problem of how our general knowledge, including our certainty about that knowedge, accumulates in the first place.

The frequency with which such plausible statements are heard seemed likely to be an important variable in this problem. Humans are profoundly sensitive to frequency (e.g., Estes, 1964, 1976; Underwood, 1971): Subjects can make reasonably accurate judgments of the frequency of events in an experiment (e.g., Hintzman, 1969); they can make frequency judgments of real world events, e.g., single words (Shapiro, 1969), single letters (Attneave, 1953), and pairs of letters (Underwood, 1971), that correlate with their actual frequency of occurrence; they can also make the rather fine grained distinction between the frequency of verbatim repetitions of sentences and of their paraphrased repetitions (Gude & Zechmeister, 1975). Differences in frequency between items will alter a subject's choice in a verbal discrimination task (Ekstrand, Wallace, & Underwood, 1966), in probability learning tasks (Estes, 1976), and will influence simple retention measures such as recognition and recall (Underwood, Zimmerman, & Freund, 1971). Frequency might also serve as the major access route that plausible statements have into our pool of general knowledge. That is, the more often you hear that 50,000 people live in Greenland, even if you do so in contexts that are explicitly ambiguous or equivocal, the more certain you will become that indeed they do. Such was the logic underlying the present experiment.

METHOD

Design

On three successive occasions, each separated by a 2-week interval, subjects heard a series of 60 plausible assertions and rated each for its validity on a 7-point scale. Twenty of the first sixty statements were selected as critical items and occurred on each of the three presentations. All other items were new. Validity ratings of the 20 repeated assertions were compared with those for nonrepeated statements. Crossed with the repeated and nonrepeated variable and the sessions variable, was a third variable, the actual truth or falsity of the assertion. The design was then a 2 (repeated vs. nonrepeated statements) × 3 (sessions) × 2 (truth vs. falsity) within-subject factorial.

Materials

A total of 140 plausible assertions were formed by culling reference works on 10 general topics: history, government and politics, sports, political science, biology and medicine, current affairs, the arts, geography, demography, and religion and social customs. Based on these references, statements were selected and written so that they were plausible without being so familiar that most students would know with certainty whether or not they were true. Fourteen instances were written for each category, seven of which were true, and seven of which were false. Two examples of statements, one true and one false, from each of the ten topics may be seen in Table 1.

From the total pool of 140 statements, 20 were selected to serve as the repeated items by choosing one true and one false statement from each of the 10 general categories. Within these constraints, the particular statements that served as repeated items were randomly selected.

The remaining 120 statements were then divided into 3 groups of 40 each, half of which were true and half of which were false. Each of the 3 groups of 40 were then combined with the 20 repeated statements to generate the 3 sets of 60 statements presented to the students for their validity judgments. Within a set of 60, statements were assigned to a position randomly with the constraint that only non-

TABLE 1

EXAMPLES OF INSTANCES FROM THE TEN KNOWLEDGE CATEGORIES

Category	Validity	Example
History	True	Kentucky was the first state west of the Alleghenies to be settled by pioneers.
	False	The People's Republic of China was founded in 1947.
Government & Politics	True	French horn players get cash bonuses to stay in the U.S. Army
	False	Zachary Taylor was the first President to die in office.
Current Affairs	True	About 1.6 billion items of litter are tossed away each year on California public lands.
	False	Total U.S. defense spending has risen steadily since 1965.
Sports	True	The Philadelphia Phillies have won only two National League pennants since 1900.
	False	Tulane defeated Columbia in the first Sugar Bowl Game.
Physical Science	True	Lithium is the lightest of all metals.
	False	It takes twice as much force to move a ton of freight by railroad as it does by truck.
Biological Science	True	The thigh bone is the longest bone in the human body.
	False	The capybara is the largest of the marsupials.
The Arts	True	Ernest Hemingway received a Pulitzer Prize for *The Old Man and the Sea.*
	False	The largest museum in the world is the Louvre in Paris.
Geography	True	Australia is approximately equal in area to the continental United States.
	False	Outside of New York and Chicago, the tallest building in America is found in Dallas.
Demography	True	Cairo, Egypt has a larger population than Chicago, Illinois.
	False	In the U.S., divorced people outnumber those who are widowed.
Religion & Custom	True	In Malaya, if a man goes to jail for being drunk, his wife goes too.
	False	Divorce is found only in technically advanced societies.

repeated items occupied the first 10 and last 10 slots in the list. Repeated and nonrepeated items were then interspersed throughout the middle 40 slots in a list. Critical comparisons on the three variables in the study were performed on the ratings assigned to items in these middle slots. The decision to use only nonrepeated statements as recency and primacy buffers was based on the following logic: The statements most likely to be remembered after a session were those in the buffer positions. If a curious and scholarly student were to look up a statement, these buffer items would be his most likely targets. That would leave us with a reasonable probability that the repeated items were not researched and so any changes in validity ratings could be attributed to the repetition manipulation rather than to extraexperimentally acquired knowledge.

The three sets of 60 statements were then tape recorded by a speaker using standard English pronunciation. They were recorded with 10 seconds between the beginning of one statement and the beginning of the next. Since on the average it took 5 seconds to complete a statement, there were approximately 5 seconds of quiet time between the end of one statement and the beginning of the next. Two different orderings of the 3 sets of 60 statements were selected with half of the subjects run in one order and half in the other.

Procedure

Subjects were instructed that they would hear statements that might or might not be true and that each one should be rated on a 7-point scale where four indicated "uncertain," five indicated "possibly true," six "probably true," and seven "definitely true." The same adjectives were used to describe false statements and occurred respectively with the ratings numbered three, two, and one. Subjects were informed that they should rate each item immediately after it occurred and prior to the next statement. They were told that they were involved in validating a new test of the general knowledge of college students.

Subjects

There were a total of 40 college student subjects whose data were used in analyzing the results. The data from 10 subjects had to be discarded because they did not attend all three sessions. Twenty of the subjects were run in small groups ranging in size from 1 to 10, while the remaining subjects were run in an intact classroom group.

RESULTS

The dependent measure was the mean rating assigned by a subject on a given session to those items in each of four categories: true items that were either repeated across sessions or not, and false items that were also either

repeated or not. While for the repeated statements 10 were true and 10 were false, this was not the case for the 20 nonrepeated statements in the critical middle 40 positions. Due to our procedure of randomly assigning nonrepeated items throughout the list, there were between 8 and 12 true statements in the critical middle positions of the three different tapes in the experiment. We judged it to be more important to eliminate response bias and curiosity factors that might influence the initial and terminal positions than to have an equal number of true and false statements in the middle segment. This was of course critical in our decision to use as a dependent measure each subject's mean rating on all those items that were a member of each of the four categories in the experiment. This enabled a straightforward comparison of the ratings given to the different types of critical items. A second consideration would also have led us to the decision to use mean ratings: Although subjects were instructed to give a rating for each item, they did not. Of the 7200 statements to be rated (60 statements × 3 sessions × 40 subjects), 29 were not. Twenty of the unrated events fell on two nonrepeated statements in the second session for the 20 subjects run in the intact classroom. This was no doubt the result of a temporary disruption. Only two events were not rated by those subjects run in small groups, a situation in which the experimenter has more control over the environment.

A preliminary analysis was conducted to determine whether there were differences in performance between the subjects run in small groups and those run in the intact classroom. There were none and this variable was collapsed across in all further analyses.

As can be seen in Table 2, the mean validity judgments ranged from a low of 4.04 to a high of 4.80. That is, they ranged from "uncertain" to "possibly true." Of course some items received much higher and others much lower ratings, but in general these ratings confirm that our item selection procedure was successful in providing statements that were plausible

TABLE 2

MEANS AND STANDARD DEVIATIONS[a] ASSIGNED TO GROUPS OF STATEMENTS ACROSS
SUCCESSIVE SESSIONS

Statement type	Session 1		Session 2		Session 3	
	M	SD	M	SD	M	SD
Repeated						
True	4.52	.60	4.80	.49	4.80	.63
False	4.18	.51	4.54	.57	4.67	.46
M	4.35		4.67		4.74	
Not repeated						
True	4.42	.52	4.30	.48	4.29	.70
False	4.09	.68	4.15	.54	4.04	.55
M	4.25		4.22		4.16	

[a] The standard deviations are based on subject means and are not means of within-subject standard deviations.

but unlikely to be in the knowledge base of most college students. These data were then analyzed by means of a 2 (repeated vs. nonrepeated) × 2 (true vs. false) × 3 (sessions) repeated measures analysis of variance.

On the average, subjects assigned higher validity ratings to statements that were true ($M = 4.52$) than to statements that were false ($M = 4.28$). The difference between true and false statements was significant, $F(1, 39) = 34.7$, $MS_e = .20$, $p < .01$. This dimension did not enter into any interactions with the remaining variables. Our students either had some reliable information on at least some of the statements for there to be a significant difference between those that were true and those that were false, or else our true statements were slightly more plausible than our false statements.

As can be seen from Table 2, the average rating assigned to repeated statements increased across successive sessions, while the rating assigned to nonrepeated statements diminished slightly. The interaction between statement repetition and session was significant, $F(2, 78) = 11.80$, $MS_e = .20$, $p < .01$. This interaction overrides the significant main effects seen for both sessions, $F(2, 78) = 6.00$, $MS_e = .19$, $p < .01$, and statement occurrence,

$F(1, 39) = 46.83$, $MS_e = .35$, $p < .01$. Simple main effects tests confirmed the apparent nature of the interaction. Validity ratings assigned to repeated items increased across successive tests, $F(2, 78) = 16.34$, $MS_e = .41$, $p < .01$, while the validity ratings assigned to nonrepeated statements did not change, $F < 1$. It is important to note that during the first session there was no difference in the ratings assigned to those statements that would later recur in sessions 2 and 3 and the ratings assigned those statements that would not recur, $F(1, 39) = 1.69$, $MS_e = .45$, $p > .05$. Thus, repeated statements are more likely to be judged as "true" than are similar, nonrepeated statements.

DISCUSSION

The present research has demonstrated that the repetition of a plausible statement increases a person's belief in the referential validity or truth of that statement. Other research has demonstrated the sensitivity of the information processing system to the frequency variable (cf., Estes, 1976; Underwood, 1971). Indeed, Underwood (1971) has proposed that frequency is the attribute of memory that underlies our ability to accurately distinguish

old events from new events. Furthermore, several recent experiments by social psychologists have indicated a relationship between frequency of exposure to a stimulus and positive affect for the stimulus (Smith & Dorfman, 1975; Stang, 1975; Zajonc, 1968). In the present experiments, the subjects' judgments that repeated statements were more probably true than nonrepeated statements occurred in a situation in which there was no verifying information available concerning the actual truth or falsity of the statements. Frequency, then, must have served as a criterion of certitude for our subjects. Indeed, the present experiment appears to lend empirical support to the idea that "if people are told something often enough, they'll believe it." In particular, it should be noted that the increase in validity ratings with repetition was equivalent for true and for false statements, despite the fact that subjects succeeded in discriminating between them. Furthermore, the increase in validity ratings occurred for an extremely diverse set of statements, which suggest that the effect of frequency upon the rated validity of statements is a general rather than a context specific phenomenon.

The precise role of frequency in cognition and memory is still unclear of course. Nevertheless, a rapidly growing body of evidence indicates that frequency is a key attribute of memory, playing a fundamental role in discriminating among memories (Underwood, 1971), in developing positive affect for a stimulus (Zajonc, 1968), and in attributing referential validity to plausible statements.

REFERENCES

ATTNEAVE, F. Psychological probability as a function of experienced frequency. *Journal of Experimental Psychology*, 1953, 46, 81–86.

BRANSFORD, J. D., & FRANKS, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, 2, 331–350.

EKSTRAND, B. R., WALLACE, W. P., & UNDERWOOD, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, 73, 566–578.

ESTES, W. K. Probability learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.

ESTES, W. K. The cognitive side of probability learning. *Psychological Review*, 1976, 83, 37–64.

GUDE, C., & ZECHMEISTER, E. G. Frequency judgments for the 'gist' of sentences. *American Journal of Psychology*, 1975, 88, 385–396.

HINTZMAN, D. L. Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, 80, 139–145.

ORNE, M. T. Communication by the total experimental situation: Why is it important, how is it evaluated, and its significance for the ecological validity of findings. In P. Pliner, L. Krames, & T. Alloway (Eds.), *Communication and affect: Language and thought*. New York: Academic Press, 1973.

ROSENTHAL, R., & ROSNOW, R. L. *The volunteer subject*. New York: Wiley-Interscience, 1975.

SHAPIRO, B. J. The subjective estimate of relative word frequency. *Journal of Verbal Learning and Verbal Behavior*, 1969, 8, 248–251.

SMITH, G. F., & DORFMAN, D. D. The effect of stimulus uncertainty on the relationship between frequency of exposure and liking. *Journal of Personality and Social Psychology*, 1975, 31, 150–155.

STANG, D. J. Effects of "mere exposure" on learning and affect. *Journal of Personality and Social Psychology*, 1975, 31, 7–12.

UNDERWOOD, B. J. Recognition memory. In H. H. Kendler & J. T. Spence (Eds.), *Essays in neobehaviorism*. New York: Appleton–Century–Crofts, 1971.

UNDERWOOD, B. J., ZIMMERMAN, J., & FREUND, J. S. Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology*, 1971, 87, 149–162.

ZAJONC, R. B. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph Supplement*, 1968, 9, (2, Part 2).