

Long-term benefits of seeding the knowledge base

NORMAN R. BROWN

University of Alberta, Edmonton, Alberta, Canada

and

ROBERT S. SIEGLER

Carnegie Mellon University, Pittsburgh, Pennsylvania

Exposure to numerical examples (seed facts) produced a substantial long-term reduction in domain-specific innumeracy. In particular, learning the populations of 24 seed countries improved accuracy of estimates of the populations of 75 untrained countries, both at the time of learning and 4 months later. Consistent with abstraction-based theories of learning and memory, the benefits of having been exposed to the seed facts were as large 4 months after the exposure as immediately after it, despite the specific populations of the seed countries having been forgotten during the interval.

Numbers often describe central properties of objects, events, experiences, and entities. Without knowledge of when events occurred, what distances separate cities and countries, and how many people live in various countries, deep understanding of history, geography and current events is impossible. Despite the importance of quantitative information in many real-world and scholastic domains, people are not particularly good at remembering numerical facts or at accurately inferring the values of quantities not stored in memory (Nickerson, 1981; Paulos, 1990; Ravitch & Finn, 1987). In this article, we investigate a method, *seeding the knowledge base*, that has considerable potential for reducing such domain-specific innumeracy (Brown & Siegler, 1993).

The seeding procedure is a simple one. Subjects are first presented with a set of items and asked to estimate the value of a particular quantitative property (e.g., a country's population). Next, they are exposed to the correct values for a subset of these items. Finally, a second set of estimates is collected. The effectiveness of the procedure is evaluated by comparing estimation accuracy for transfer (i.e., "nonseed") items across the two tests. The procedure has proved useful for improving estimates of national populations and land areas, city-to-city distances, and latitudes and longitudes of geographic entities (Brown & Siegler, 1993, 1996).

These findings raise interesting practical and theoretical issues. On a practical level, it is important to know

whether the benefits produced by the seeding procedure are long lasting. If they are, it should be possible to improve quantitative understanding in many substantive areas by teaching students a relatively small set of well-chosen numerical examples.

On a theoretical level, the interest lies in distinguishing between two explanations for the seeding effects: an *anchoring* explanation and a *feedback/induction* explanation. The anchoring explanation suggests that seeding improves estimation by providing specific numerical reference points. In contrast, the feedback/induction explanation suggests that the procedure works because seed facts allow subjects to evaluate and adjust their knowledge of the statistical properties of the relevant quantitative dimension.

In the present study, we adopted a longitudinal approach to addressing these practical and theoretical issues. During an initial session, subjects estimated the population of 99 countries. Next, they learned the actual populations of 24 of the countries. They then again estimated populations of all 99 countries. Four months later, subjects returned to the laboratory and produced a third set of estimates for all 99 countries.

This design allowed us to observe both immediate and long-term effects of the seeding procedure. If seeding produces a long-term change in subjects' understanding of the populations, then follow-up estimates for transfer countries should be more accurate than pretest estimates. We were also able to track knowledge of seed facts over time and to assess the relation between subjects' memory for these facts and the accuracy of their transfer estimates. These data were of particular interest, because the anchoring explanation implies that memory for seed facts should be closely related to performance on transfer items, whereas the feedback/induction explanation does not.

The anchoring position is based on the view that real-world estimates are often generated by recalling poten-

This research was supported in part by an NSERC operating grant and an NIMH postdoctoral fellowship awarded to the first author, and in part by a Mellon Foundation grant awarded to the second author. The authors thank Jeff Bisanz, Fred Conrad, and Monica Dashen for their comments on a draft of this article, and Cathy Dennler for her assistance in running the study. Correspondence should be addressed to N. R. Brown, Department of Psychology, University of Alberta, Edmonton, AB, Canada, T6G 2E9 (e-mail: nbrown@psych.ualberta.ca).

tially relevant facts and using them to draw plausible inferences (Collins & Michalski, 1989). People often reconstruct dates for personal events (e.g., Burt, 1992; Thompson, 1982) and public ones (e.g., Brown, 1990; Friedman & Wilkins, 1985). Estimates for other real-world quantities also may be reconstructed. For example, a person estimating the population of Poland might recall that France has a population of about 60 million, that Romania has a population of about 20 million, and that Poland's population is in between that of France and Romania, and infer from these facts that Poland's population is between 20 million and 60 million. As this example illustrates, more than one fact or inference will often be necessary to generate an estimate. As a result, estimates should improve when people have access to many numerical facts. From this perspective, the seeding procedure works because it increases the number and availability of potentially relevant quantitative facts, thus allowing subjects to generate more and better quantitative inferences. Conversely, reduced access to these facts should produce a corresponding decrease in estimation accuracy. Thus, the anchoring position implies that if subjects forget seed facts over the 4-month retention interval, follow-up transfer estimates should be less accurate than estimates made immediately after the facts were learned.

The feedback/induction position de-emphasizes storage and retrieval of specific numerical examples. Instead, it assumes that *metric* beliefs (i.e., knowledge of the statistical properties of the target dimension—e.g., mean, range, variance, distributional form, etc.) are critical in real-world estimation (Brown & Siegler, 1993). In this view, people first estimate the relative size of the entity in question and then select a value from the corresponding portion of distribution of quantitative values (e.g., "Poland has a medium sized population. Midsize countries have populations of about 40 million. Therefore, Poland has a population of about 40 million."). From this perspective, the seeding procedure is effective because seed facts provide both feedback on the accuracy of relevant metric beliefs and a basis for revising these beliefs when they are mistaken.

Within this feedback/induction position, after people update their task-relevant metric beliefs, they no longer need to remember the seed facts. As a result, memory for seed facts should be unrelated to accuracy of transfer estimates. Thus, the feedback/induction position predicts that posttest and follow-up transfer estimates should be equally accurate, even if the seed facts are forgotten over the 4-month interval.

Although the present study was primarily designed to address issues concerning the seeding procedures, its implications can be cast in broader terms. The anchoring and feedback/induction positions map onto two very different approaches to learning and memory. Like instance-based learning theories, the anchoring position assumes that specific instances are stored in memory and that these instances are later retrieved and used to guide higher level cognitive processes (e.g., Hintzman, 1986; Jacoby &

Brooks, 1984; Logan, 1988; Medin & Schaffer, 1978; Ross, 1984). In contrast, the feedback/induction position, with its emphasis on the creation, storage, and application of general knowledge, resembles abstraction-based or prototype theories (e.g., Posner & Keele, 1970; Rosch, 1975; Smith & Medin, 1981). Seen from this perspective, the present research can help determine whether an instance-based approach is sufficient to account for the effect of numerical examples on real-world estimation, or whether it is necessary to propose that people also form abstract quantitative beliefs and use them in the estimation process.

METHOD

Materials and Procedure

Data were collected during two sessions, 4 months apart. In the initial session, four tasks were presented. First, the subjects provided knowledge ratings for 99 countries, using a 0 (*no knowledge*) to 9 (*a great deal of knowledge*) scale. Second, they estimated each country's current population. Third, they learned the actual populations of 24 seed countries. During the learning phase, the subjects studied and were tested on each seed country's population four times. Finally, the subjects provided a second set of estimates for all 99 countries (including the 24 seed countries). In the follow-up sessions, the subjects again estimated the populations of all 99 countries. Country names always were presented one at a time, in a unique random order, at the center of a computer-controlled video display; all responses were typed at the computer's keyboard.

The 99 test countries represented all but one of the countries that had populations of at least 4 million in 1989 (*Information Please Almanac*, 1989). The one exception was the United States, whose population was given to the subjects as an example before the first estimation task. Data collected in prior experiments were used to select the seed set. This set was constructed so that there were 6 countries in each of the four cells created by crossing country knowledge (high vs. low) and estimation accuracy (high vs. low). See Brown and Siegler (1993) for further details.

Subjects

Twenty-five Carnegie Mellon undergraduates were given course credit for participating in the first session, which was conducted in December 1989. Four months later (April 1990), all subjects were contacted by telephone and offered \$5 to participate in an unspecified psychology experiment. Twenty of the original 25 participated.

Results

Two measures of estimation accuracy are reported below. One of these, order of magnitude error (OME), was computed for each response according to the following formula:

$$\text{OME} = |\log_{10}(\text{estimate value}/\text{actual value})|.$$

OME converts estimation errors to a percentage of an order of magnitude. It was used because the stimulus scale spans several orders of magnitude, and, as a result, a given response could be orders of magnitude too large or too small. Therefore, it seemed appropriate to treat order of magnitude as the relevant unit of measurement (Brown & Siegler, 1992; Nickerson, 1981).

For each of the 20 returning subjects and each estimation task (pretest, posttest, and follow-up), separate means were computed for the 24 seed-country OME and the 75

transfer-country OME, yielding 6 means per subject. For each subject, a rank-order correlation between estimated and actual population was also computed for both seed and transfer countries, for each estimation task. These correlations provided an index of the subjects' knowledge of relative country size.

Mean OMEs and (r -to- z transformed) rank-order correlations were submitted to separate 2 (country type—seed vs. transfer) \times 3 (test—pretest vs. posttest vs. follow-up) analyses of variance. The country type \times test interaction was highly significant for both OMEs [$F(2,38) = 82.17$, $MS_e = 0.019$], and r -to- z transformed correlations [$F(2,38) = 44.95$, $MS_e = 0.089$], as were the main effects of country type and test ($p < .0001$, in all cases). OME means and the (back-transformed) mean rank-order correlations are presented in Figure 1. Least significant differences tests ($p < .01$) were used to test differences between the means.

The interactions illustrated in Figure 1A indicate that the seeding procedure produced long-term improvements in estimation accuracy for the transfer countries, despite memory for the seed facts fading over the 4-month retention interval. The OME means for the transfer estimates provide a justification for the former claim (see Figure 1A). Follow-up estimates for transfer countries were 37% more accurate than pretest estimates for them (OMEs of .42 vs. .65). Of particular interest, the posttest and the follow-up OMEs (.42 and .41) were almost identical. The pretest mean differed significantly

from both the posttest and the follow-up means, which did not differ from one another.

Although the seeding procedure strongly influenced absolute accuracy, as measured by transfer OME, it did not affect relative accuracy, as measured by the transfer correlations (Figure 1B). This is consistent with the notion that population estimation involves independent metric and mapping components (Brown & Siegler, 1993) and indicates that learning the current seed set affected only the former.

Changes over time in both the OMEs and rank-order correlations indicate that the subjects learned the seed facts during the learning session but forgot them over the retention interval. In both cases, the posttest estimates were far more accurate than either the pretest estimates (evidence for learning) or the follow-up estimates (evidence for forgetting). Mean posttest OME for the seed countries (.08) reflected significantly more accurate estimates than either the pretest (.56) or follow-up (.34). Similarly, the mean posttest correlation (.88) was significantly larger than either the mean pretest (.34) or follow-up (.47) correlation.

As with the transfer countries, the difference between pretest and follow-up OMEs for the seed countries was significant, and the difference between pretest and follow-up correlations was not. Moreover, although both measures indicate that immediate posttest estimates for seed countries were significantly more accurate than immediate posttest estimates for transfer countries, estimates for seed countries were not reliably better than estimates

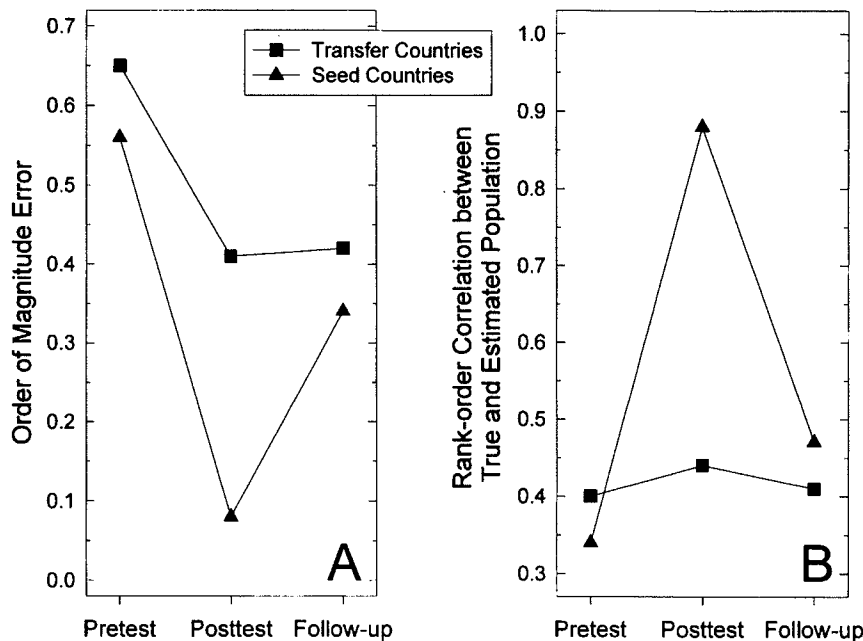


Figure 1. Mean order of magnitude error (A) and mean (back-transformed) rank-order correlation between estimated and actual population (B) computed over responses to seed and transfer countries, for the pretest, posttest, and 4-month follow-up test.

for transfer countries during the follow-up task. Thus, on the posttest, the subjects appear to have generated populations for the seed countries by recalling the facts they had just studied. In contrast, for the transfer countries on both the posttest and follow-up and for the seed countries on the follow-up, their estimates appear to have been based on mapping their perception of each country's relative size onto the distribution of population values abstracted from the seed facts.

CONCLUSION

The present results indicate that benefits of seeding the knowledge base are long lasting. For transfer countries, estimates collected 4 months after exposure to seed facts were no less accurate than estimates produced immediately after the seeding procedure, and both were far more accurate than pretest estimates. This strengthens our belief that seeding provides a simple, effective method to remedy domain-specific innumeracy.

We also found that the subjects tended to forget the seed facts during the 4-month retention interval, but that this did not affect their ability to provide accurate estimates for the transfer countries. These results are consistent with the feedback/induction position and incompatible with the anchoring position. It seems that the subjects used seed facts to evaluate and adjust their metric beliefs about national populations. These updated beliefs appear to have been maintained as part of the subject's knowledge and to have played an important role in determining the magnitude of the follow-up estimates.

Above, we suggested that the anchoring interpretation is related to instance-based theories of learning and memory and that the feedback/induction position has much in common with theories that emphasize abstraction and generalization. To the extent that these comparisons are valid, the present research suggests that longitudinal designs—ones that allow time for specific memories to fade—may be useful in distinguishing between instance-based and abstraction-based models. In general, instance-based models predict a direct relation between memory for instances and transfer performance, whereas abstraction-based positions do not. Thus, tracking instance memory and transfer performance across a reasonably long retention interval provides a way of discriminating between the alternatives (also see Brooks, Norman, & Allen, 1991).

In addition to highlighting the potential value of longitudinal designs for distinguishing between competing learning theories, the present findings support the notion that generalization plays an important role in memory and learning, at least in part because general knowledge is more resistant to forgetting than is specific knowledge. Certainly, a (well-rehearsed, richly elaborated) generalization summarizing information implicit in a set of examples should be more accessible than the individual examples that gave rise to it. This should be particularly true in situations such as the present one, where relevant examples are rarely encountered and/or imperfectly learned. Thus, we are not arguing that specific information does not exist in memory nor that specifics play no role in real-world estimation—clearly, both claims are too strong. Nonetheless, it does seem that a close, persisting relation between instance memory and transfer performance should be

demonstrated before one accepts an instance-based explanation or rejects an abstraction-based one.

REFERENCES

- BROOKS, L. R., NORMAN, G. R., & ALLEN, S. W. (1991). Role of specific similarity in a medical diagnostic task. *Journal of Experimental Psychology: General*, **120**, 278-287.
- BROWN, N. R. (1990). Organization of public events in long-term memory. *Journal of Experimental Psychology: General*, **119**, 297-314.
- BROWN, N. R., & SIEGLER, R. S. (1992). The role of availability in the estimation of national populations. *Memory & Cognition*, **20**, 406-412.
- BROWN, N. R., & SIEGLER, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review*, **100**, 511-534.
- BROWN, N. R., & SIEGLER, R. S. (1996). *Inducing statistical properties of real-world domains*. Manuscript in preparation.
- BURT, C. D. B. (1992). Retrieval characteristics of autobiographical memories: Event and date information. *Applied Cognitive Psychology*, **6**, 389-404.
- COLLINS, A. M., & MICHALSKI, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science*, **13**, 1-49.
- FRIEDMAN, W. J., & WILKINS, A. J. (1985). Scale effects in memory for time of events. *Memory & Cognition*, **13**, 168-175.
- HINTZMAN, G. L. (1986). "Schema abstraction" in a multiple-trace model. *Psychological Review*, **93**, 411-428.
- Information please almanac* (42nd ed.) (1989). Boston: Houghton Mifflin.
- JACOBY, L. L., & BROOKS, L. R. (1984). Non-analytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18, pp. 1-47). New York: Academic Press.
- LOGAN, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, **95**, 492-527.
- MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.
- NICKERSON, R. (1981). Motivated retrieval from archival memory. In J. H. Flowers (Ed.), *Nebraska Symposium on Motivation 1980* (pp. 73-119). Lincoln: University of Nebraska Press.
- PAULOS, J. A. (1990). *Innumeracy: Mathematical illiteracy and its consequences*. New York: Vantage.
- POSNER, M. I., & KEELE, S. W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, **83**, 304-308.
- RAVITCH, D., & FINN, C. E. (1987). *What do our 17-year-olds know?* New York: Harper & Row.
- ROSCHE, E. (1975). Cognitive representation of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-233.
- ROSS, B. H. (1984). Reminders and their effects on learning a cognitive skill. *Cognitive Psychology*, **16**, 371-416.
- SMITH, E. E., & MEDIN, D. L. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.
- THOMPSON, C. P. (1982). Memory for unique personal events: The roommate study. *Memory & Cognition*, **10**, 324-332.

(Manuscript received December 18, 1995;
accepted for publication March 1, 1996.)