

REAL-WORLD ESTIMATION: ESTIMATION MODES AND SEEDING EFFECTS

Norman R. Brown

I. Introduction

More people live in Ethiopia (64 million) than in the United Kingdom (59 million). Located at 34° north latitude, Atlanta, Georgia is just one degree north of Tijuana Mexico. At the moment, Bill Gates is worth about \$64 billion; the GDP of Tunisia (with its population of 9.7 million people) was \$63 billion in 2000. Montreal is about 2900 kilometers from Edmonton, Alberta; the distance between Edmonton and the city of Chihuahua in Mexico is about 2800 kilometers. At \$48,000 a new Honda S2000 cost more than a new BMW Z3 sports car which goes for \$45,000, etc, etc.

To many people, these facts taken separately, or in tandem, are surprising, even counterintuitive. From a cognitive perspective, this is an interesting response because it suggests that people do have intuitions about many different real-world quantities and that these intuitions can be very wrong. At the same time, exposure to numerical facts like these can be highly informative. This latter point has been made a number of times using a method called *seeding the knowledge-base* (Bostrom & Brown, 2001; Brown, 2001; Brown, Friedman, & Lee, 2001; Brown & Siegler, 1993, 1996, 2001; Friedman & Brown, 2000a,b; Friedman, Kerkman, & Brown, in press-b; LaVoie, Bourne & Healy, in press; Murray & Brown, 2001; Walbaum, 1997). In the usual seeding experiment, participants first provide numerical estimates for a set of items; they then learn the actual values of a subset

of these items; and finally they provide a second set of estimates for both the seed facts and the transfer items. Typically, postseeding estimates for the transfer items are more accurate than the preseeding estimates, and at times, the seeding procedure produces dramatic reductions in estimation error.

To date, experiments that have used the standard 3-phase seeding procedure have demonstrated robust and positive seeding effects on estimates of national populations (Brown & Siegler, 1993, 1996, 2001; LaVoie et al., in press). Seeding has also been shown to increase estimation accuracy when people estimate latitudes and longitudes (Friedman & Brown, 2000a,b; Friedman et al., in press-b), city-to-city distances (Brown & Siegler, 2001; Brown et al., 2001), automobile prices (Murray & Brown, 2001), fatality rates (Bostrom & Brown, 2001), university tuitions (Lawson & Bhagat, in press), and the nutritional value of fast food (Walbaum, 1997). In addition, recent studies that employed a simplified procedure (one seed fact and one transfer item) have shown that seeding also improves accuracy when people estimate corporate sales figures, CD sales, national land area, populations of endangered species, and the heights of mountains (Beck & Carlson, 1998; Brown, 2001; Friel & Carlson, 2000). Finally, there is evidence that seeding effects can be long-lived. For example, Brown and Siegler (1996) found that the benefits of seeding were undiminished after 4 months (also see LaVoie et al., in press).

Many of the seeding studies just cited are part of a broader research program, one that is concerned with understanding (a) the processes used to generate real-world estimates, (b) the nature and representation of the domain-specific knowledge that these processes operate on, and (c) the ways that seed facts and other forms of potentially relevant information affect these processes and this knowledge. This work has been motivated by both theoretical and practical concerns. On a theoretical level, this project represents an attempt to understand a little-studied but important form of thinking, a form that might be called *open-ended complex cognition*. Although experimental methods are used to study real-world estimation, there are fundamental differences between this research program and the experimental study of related topics such as problem solving, reasoning, and judgment and decision making. Specifically, the mainstream study of higher-level cognition has relied heavily on knowledge-lean tasks and has been primarily concerned with understanding the application of a small set of formal operations to a well-defined problem space and/or identifying and explicating deviations from normatively prescribed patterns of performance. In contrast, the central goal of current research programs is to develop an empirically grounded understanding of the nature and use of domain-specific knowledge.

More generally, this research has been motivated by the observations that people have a great deal of real-world knowledge (i.e., facts about the world learned through experience), that they continually and successfully draw on this knowledge to understand the world and to behave intelligently in it, and that the knowledge used to accomplish these feats is often incomplete, and at times inaccurate and

contradictory. These observations raise two broad issues, the first having to do with acquisition, representation, and organization of real-world knowledge, and the second having to do with the utilization of this knowledge. It turns out that the study of real-world estimation sheds light on sets of both issues. On the one hand, quantitative estimates are affected by and often reflect the content and structure of the information used to generate them. As a result, quantitative estimates can reveal much about a given domain of knowledge (Brown, 1990; Brown & Siegler, 1993; Friedman & Brown, 2000a,b; Friedman, Brown, & McGaffey, in press-a; Murray & Brown, 2001). On the other hand, because real-world knowledge is complex, and relevant between-domain differences are large, estimation strategies vary from task to task, and different strategies often compete within the same task. It follows that the study of real-world estimation should produce information about the range of estimation strategies people use, the factors that influence strategy selection, and the ways that people coordinate competing sources of information (Brown, 1995, 1997; Brown & Siegler, 1993; Conrad, Brown, & Cashman, 1998). In brief, because performance on many real-world estimation tasks reflects the systematic interplay of process and content, it is possible to employ these tasks to investigate both.

In addition to its theoretical contribution, an understanding of real-world estimation has obvious practical value. Survey methodology is one area where this sort of research could have immediate impact. Here, an accurate model of real-world estimation should provide guidance for the designers of questionnaires intended to elicit quantitative information from survey respondents and should also aid in the interpretation of biased or inconsistent survey responses (Brown & Sinclair, 1999; Conrad et al., 1998; Sudman, Bradburn, & Schwarz, 1996). There are also direct educational applications of this work. For example, the evidence that seeding often improves estimation accuracy suggests that educators should be able to counter domain-specific innumeracy by exposing their students to a few key facts. However, there is also evidence that the effectiveness of the seeding procedure depends on what people already know about the target domain and on the specific identities of the seeds and the transfer items (Brown & Siegler, 1993; Friedman & Brown, 2000a,b; Kerkman, Friedman, Brown, Stea, & McCormick, 2001; Murray & Brown, 2001). Thus, the selection of an optimal set of seed facts may well require a thorough understanding of the relevant domain knowledge and of those cognitive mechanisms that produce positive seeding effects.

The goal of this chapter is to provide a synthetic overview of the research on real-world estimation and seeding. To this end, I adopt a *plausible reasoning* architecture as a general processing framework (Collins & Michalski, 1989). I identify two basic estimation modes, the *numerical-retrieval* mode and the *ordinal-conversion* mode, and describe a number of processes engaged by each of these. In addition, I define two types of seeding effects, one of which involves *revising the metric*, and the other (*re*)*partitioning the range*. In the course of developing these points, I also contend that ordinal conversion is far more common than

numerical reconstruction, that people generally use the same mode to generate their preseeded estimates and their postseeded estimates, and that seeds are effective because they allow people to identify and revise inaccurate assumptions about the magnitude and/or the division of the response range.

II. Estimation Processes

A. PLAUSIBLE REASONING AND THE RETRIEVAL–INFERENCE CYCLE

In a real-world estimation task, experimental participants are presented with a set of target items and a target dimension and are required to estimate as accurately as possible the value of each target item. For example, participants might be presented with a set of country names (target items) and be asked to estimate the current population (the target dimension) of each. Or, they might be given names of a set of automobiles and be required to estimate the current sales price of each, etc. One way to respond to such questions is to retrieve the value directly from memory and state it as an answer. However, the obvious precondition for providing this type of response is rarely met; in most domains, prestored numerical facts are few and far between (Nickerson, 1980; Paulos, 1990). Although people rarely *know* the answers to questions posed in these tasks, they often produce responses that are at least partially correct; it is common to find a reliable correlation between the magnitude of the numerical responses and the actual magnitudes of the to-be-estimated values, and, under some conditions, the numerical values provided by participants are quite close to (but different from) the to-be-estimated values.

The dearth of accessible numerical facts rules out simple fact retrieval as a common response strategy, and the existence of a nonarbitrary relation between the subjective and objective values indicates that people generally do not answer these questions by guessing at random. Rather, people often find a way to use their limited knowledge of the world to arrive at a fairly sound estimate, or at least to produce a response that bears some correspondence to the to-be-estimated value. In the next section, I discuss two estimation modes that describe how these educated guesses are produced. Before moving on to this discussion, however, it is useful to make explicit several of the assumptions that underpin the current approach.

The first of these assumptions is that a *retrieval–inference cycle* is almost invariably engaged when people generate real-world estimates. This cycle consists of three processes carried in sequence and often iteratively (Brown, 1990; Norman & Bobrow, 1979; Williams & Hollan, 1981). The first of these is a retrieval process, the second an inference process, and the third an evaluative process. On this view, when a question is posed about a target item, information about that item is accessed in memory. If the answer to the question cannot be recalled, a fact related to the target item and the target value is retrieved. If this retrieved fact is

relevant to the question at hand, it triggers a *plausible inference* which may serve to narrow the range of possible responses or to increase the certainty associated with a particular response. Following this inference, an evaluation is performed to determine whether task demands have been met. Often, a single inference will not provide sufficient constraint to allow for a precise confident estimate of the target value. As a result, people may iterate through this cycle, retrieving information and drawing inferences until they conclude that they have come up with a good answer or until they run out of time or potentially relevant information.

For example, a person attempting to determine when Desert Storm took place might recall that the first George Bush was president at the time and that Bush's term of office ran from 1988 to 1992. Together these facts allow for the confident inference that Desert Storm took place during the late 1980s or early 1990s. Suppose during the second cycle, this individual recalled a discussion with a high-school friend and that she or he attended high school between 1990 and 1994. This fact places a new, more restrictive, lower bound on the range, but leaves the upper bound unchanged. The retrieval-inference cycle might terminate at this point and the midpoint of the 1990-to-1992 range might then be selected as an estimate. If so, this process would produce January 1991 as a response. To take another example, a person attempting to estimate the number of islands in the Indonesian archipelago might recognize that Indonesia is an island nation and recall that the state of Hawaii is composed of 8 islands. Given these two facts and assuming that countries are generally larger than states, this person might conclude there are, say 20 islands in Indonesia. It is worth noting that these examples demonstrate that plausible reasoning can produce both accurate and inaccurate responses: Allied troops invaded Iraq in February, 1991; the Indonesian Archipelago consists of some 15,000 islands.

Collins and Michalski (1989) have demonstrated that it is possible to develop a detailed and formal representation of plausible inference patterns, and other researchers have argued that people take this type of "problem solving" or "reconstructive" approach when they are required to recall autobiographical events or facts in response to specific cues (Brown & Schopflocher, 1998; Conway, 1996; Reisser, Black, & Abelson, 1985; Williams & Hollan, 1981). It appears that this general approach works as well as it does because the information most relevant to the target item and the current task is often recalled by the retrieval process and because plausible inferences allow people to take advantage of the many (quasi-) logical implications that are sanctioned by their knowledge and beliefs without requiring them to encode and maintain massive quantities of detailed information. In addition, the open-ended nature of the retrieval-inference cycle means that a given conclusion may be corroborated by multiple lines of evidence.

A second related assumption is that people take an *opportunistic* approach to real-world estimation. In other words, they generally do not approach these questions with a deliberate response strategy in mind. Rather, it appears that people

“blindly” work toward a response using readily accessible knowledge to support a series of plausible inferences. Thus, the strategy that is adopted and the particular sequence of steps involved in producing a response are determined by the facts that happen to be retrieved. And, the availability of a given fact at a given time is determined by how strongly it is associated with the target item, and how recently it has been accessed. One implication of this view is that these retrieved facts should provide a representative sample of knowledge and that this sample should reflect the organization of this information in memory. This implication has proved to be valid at least for one type of knowledge, the knowledge that people have of public events (Brown, 1990).

One final point. To date, the strongest evidence that plausible reasoning plays a role in real-world estimation comes from the work on event dating (see below). This task may be a special one because event knowledge is often embedded in a rich network of interevent associations (Brown, 1990) and because temporal landmarks are relatively common (Shum, 1998). Nonetheless, informal protocols suggest that the use of plausible reasoning is not restricted to date estimation nor does it depend on access to numerical facts. Rather, the retrieval–inference cycle that drives plausible reasoning can be evoked for a number of different reasons. At least in principle, plausible reasoning can be used when people transform a set of vague metric beliefs into a functional response range; when they partition a derived range or an experimenter-provided range; when they assign an ordinal value to a target item in the absence of numerical facts; when they verify or fine-tune familiarity-based intuitions; and when they guess in a systemic manner. It is true that reconstructive processing observed in date estimation provides an excellent example of plausible reasoning in a real-world estimation task, but there are many others.

B. ESTIMATION MODES

When people estimate real-world quantities, they access relevant information from memory and draw plausible inferences. It has been possible to move beyond this very general claim and develop a more detailed characterization of real-world estimation, one that identifies two very different estimation modes, and to identify some of their underlying processes (see Fig. 1). These two modes, *numerical retrieval* and *ordinal conversion*, are discussed in detail in the next two sections. Briefly, an estimate involves numerical retrieval when at least one relevant numerical fact is recalled and that fact serves as the basis for a response. In contrast, ordinal conversion typically involves a preparatory stage, called *setting the metric*, during which a response range is defined. Once the range has been established, estimates are generated by determining the relative or ordinal value of the target item and selecting a numerical value from the appropriate portion of the range.

There are two fundamental differences between the estimation modes. First, a precondition for the use of numerical retrieval is access to relevant domain-specific

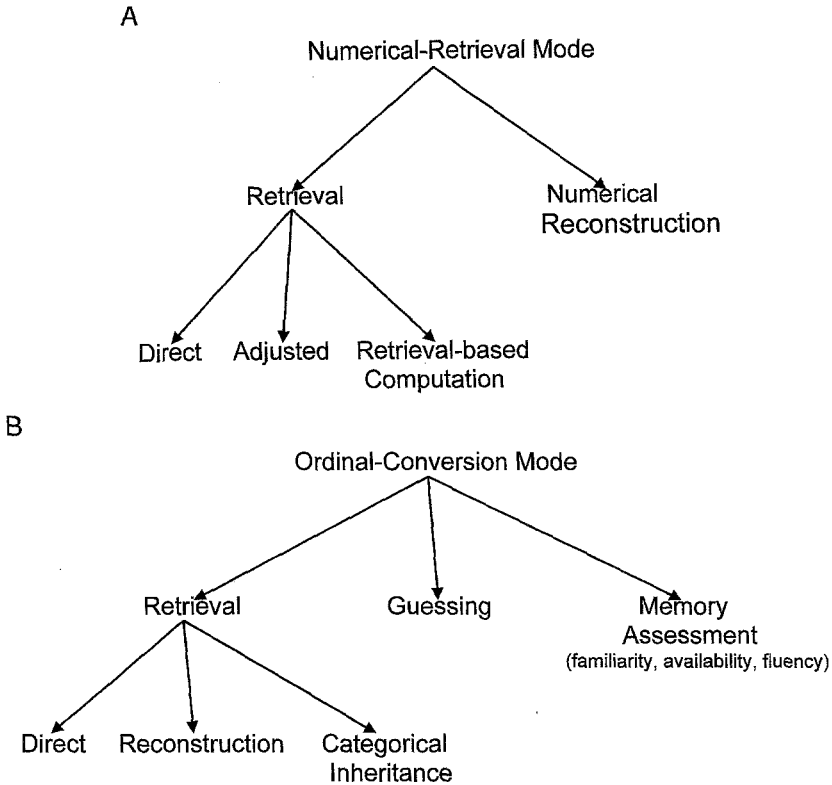


Fig. 1. A taxonomized list of processes associated with the numerical-retrieval mode (Panel A) and the ordinal-conversion mode (Panel B).

numerical facts, where a numerical fact is defined as preexisting knowledge of the numerical value posed by a particular item for a given dimension (i.e., D-Day took place on June 6, 1944; about 7.5 million people live in New York City; Boston is about 2,500 miles from LA). In contrast, there are no preconditions for use of ordinal conversion; this makes ordinal conversion the default and numerical retrieval the special case. Second, although the process of setting the metric is essential to the ordinal conversion, it appears to play no role when people rely on the various numerical retrieval processes. This is because the retrieved numerical facts are necessarily defined in metric terms and thus they bring the metric with them.

1. The Numerical Retrieval Mode

Figure 1 lists 4 retrieval-based processes. Three of these, *direct retrieval*, *adjusted retrieval*, and *retrieval-based computation*, can be used only when a relevant

numerical fact is associated with the target item. The fourth retrieval-based processes is *numerical reconstruction*. This process requires only that there be a known or inference relation between the target item and a retrieved numerical fact. Because a single fact may be related or relatable to multiple items, one would expect reconstruction to be the most common of these.

a. Fact Retrieval Direct retrieval is the simplest of the retrieval processes. When the to-be-estimated value has already been stored in memory, a response can be made simply by recalling and stating the recovered value. For example, a person asked to estimate the distance between New York and Chicago might simply know and state that the two cities are about 700 miles apart. Of course, if the value of the to-be-estimated quantity is known beforehand, the response is not an estimate but a statement of fact. However, because numerical facts are quite scarce, this strategy is uncommon.

An estimate involves adjusted retrieval when a person has succeeded in retrieving an appropriate numerical value associated with the target item, but he or she is also aware that the value is likely to be wrong for some reason. An example of this would be someone who is attempting to estimate the current population of Los Angeles and who recalls that it had a population of about 3.5 million in 1990. Given that the LA population has grown since then, this value is bound to be too small, so it would be necessary to compensate for this by adjusting the estimate upward.

The third numerical process, retrieval-based computation, also assumes that a numerical value is associated with the target item and is retrieved during the estimation task. However, in this case, the retrieved value must be mathematically transformed prior to responding. Consider a sports fan who is trying to determine the attendance at last night's well-attended home game. This person might know that the local basketball arena holds 18,500 and might figure that 75% of the seats are sold on a good night. Given this fact and this belief, it would be reasonable to place attendance at about 14,000 ($\approx 18,500 * 0.75$).

At present, it appears that these retrieval strategies are quite uncommon. This is not surprising given the dearth of explicit numerical information in many real-world domains. It is worth noting, however, that strategies like these are often used when people estimate event frequencies. For example, when Brown and Sinclair (1999) asked several hundred undergraduates to estimate the number of sexual partners they have had in their lifetimes, about a third of the sexual sample responded with a retrieved tally—this would be classified as direct retrieval in the present scheme. Similarly, several researchers have reported that people often estimate the frequency of recurring events by retrieving a rate from memory and multiplying it by the appropriate value—this would be classified as retrieval-based computation (Blair & Burton, 1987; Conrad et al, 1998; Menon, 1993). Finally, an analogue to adjusted retrieval has been seen in the lab. Here it is common for participants to

generate frequency estimates by retrieving some subset of relevant instances and then adjusting their enumerated counts to account for the unretrieved items (Brown, 1995). Thus, although numerical strategies appear to be uncommon in many real-world estimation tasks, people can and do retrieve, adjust, and transform numerical values when they are available.

b. Numerical Reconstruction The three retrieval strategies can only be used when a relevant numerical fact has been stored with the target item. Because this situation is uncommon, people are rarely in the position of using one of the retrieval strategies to generate their estimates. Of course, numerical facts can be very useful when there is a known or inferable relation between them and a target item.

Examples of numerical reconstruction were presented above as part of the discussion on plausible reasoning and the retrieval inference cycle (also see Brown 1990, Experiment 1). Certainly, numerical reconstruction fits easily within this framework. Basically, this strategy is used when the retrieval process has access to numerical facts and these facts are used to support inferences that restrict the range of possible responses or to numerical reference points. There is good evidence that numerical reconstruction is common when people estimate the dates of personal and public events (Brown, 1990; Friedman, 1993; Thompson, Skowronski, Larsen, & Betz, 1996). This evidence comes in two forms. First, when people are asked to think aloud while estimating dates they often mention landmark events or temporally bound historical or personal periods (Brown, 1990). Similarly, post hoc strategy reports indicate that people rely on temporal landmarks and dated period boundaries when they date events (Thompson, Skowronski, & Betz, 1993). Second, better-known events tend to be dated more accurately than less well-known events of the same objective age (Brown, 1990; Burt, 1992). This is what one would expect if ability to access and utilize related numerical reference points and other types of temporally informative facts is related to the richness of the information associated with the target items.

Not all temporal inferences depend on the retrieval of landmark dates. As Thompson and his colleagues (1993, 1996) have pointed out, people have a good deal of knowledge about temporal regularities that they can draw on when dating personal and public events. For example, if one is attempting to determine when a particular picnic occurred, it seems reasonable to assume, based on our knowledge of the prototypical picnic, that it happened on a weekend during the summer months. Conversely, although reconstruction has been studied primarily by researchers interested in date estimation, this does not mean that dates are the only values that can be numerically reconstructed. Consider, for example, how a person from Edmonton, Alberta might estimate the population of Winnipeg, Manitoba. This individual might "reconstruct" its population by recalling that Edmonton, Alberta has about 650,000 people and that Edmonton is larger than Winnipeg. Given these two facts, one a specific numerical value associated with an item

other than the target and the second a piece of relational information that links the numerical fact to the target item, it would be reasonable to conclude that Winnipeg has a population of about 600,000. Note that although our hypothetical Edmontorian did not know the population of Winnipeg, his/her estimate was a good one—Winnipeg's current population is 570,000.

At present, we do not know how widespread numerical reconstruction is. Generalizing from the event dating situation, one might predict that reconstruction should be common when the numerical facts are relatively common and when there is a fairly rich network of relevant interitem associations. A dearth of numerical reference points would hamper reconstruction because it would limit the number and precision of the numerical inferences that could be drawn and because a given reference point is unlikely to have a known or inferable relation to all target items. Similarly, a sparse network of interitem associations would hamper access to and interpretation of relevant numerical facts.

2. *The Ordinal-Conversion Strategy*

The presence of numerical facts and the existence of a reasonably rich knowledge domain appear to be preconditions for the use of numerical strategies. In this section, I describe the estimation mode that people can rely on to produce real-world estimates when these conditions are not met.

Several years ago, Siegler and I proposed that people depend on two types of knowledge when they generate real-world estimates: *metric* knowledge and *mapping* knowledge (Brown & Siegler, 1993). Metric knowledge was defined as knowledge of or beliefs about pertinent statistical properties of the target dimension. At the most general level, metric beliefs provide a rough specification of the magnitudes covered by the target dimension. At a more detailed level, metric knowledge may also be used to define the width of the response range (i.e., to specify the upper and lower bound) and to partition the range in a way that reflects beliefs about its central tendency and distribution of the items across its range. Mapping knowledge refers to information that allows people to determine the relative magnitude of a given target item.

The ordinal-conversion strategy was an implicit component in the metrics-and-mapping formulation. As its name implies, this strategy consists of two core processes: an ordinal component, which determines the relative magnitude of the target item, and a conversion component, which generates a numerical response by selecting a value from the appropriate portion of the response range. Of course, to do this, it is necessary to have defined and partitioned a response range. Thus, the use of an ordinal-conversion strategy implies the existence of a prior process, one I refer to as setting the metric.

As an example of how these processes work together, consider the task that confronted the university students who took part in one of our population estimation

experiments. At the outset of an experiment, participants were informed of the current population of the United States (250 million at the time) and that the test countries were among the largest countries in the world. Because most undergraduates are not experts in demography, it is safe to assume that these people have only a very limited store of additional numerical facts; they might have known that there are more than a billion people in China and little else. Knowledge of China's population is useful though, as it provides an upper bound for the response range. But, what is the lower bound? Well, it would be less than 250 million; but how much less? Let us assume that this participant simply selected a round number, say 1 million, as the lower bound of the range.

At this point, a response range has been established that spans over 3 orders of magnitude and that includes only two reference points, the population of the United States and the population of China. Next (or within the first few trials) the participant must partition this response range. This could be done in a number of ways. For the sake of this example, let us assume that 5 categories were created: very small, small, medium, large, and very large. The population of the United States could serve to separate the large from very large populations. The next partition, separating the large populations from mid-sized populations might be set at 100 million, and the next two, separating mid-sized from small, and small from very small, might be set at 50 million and 25 million, respectively.

Now consider what would happen when Germany is presented as a test item. The absence of a rich network of relevant quantitative facts rules out numerical reconstruction. As a result, it would be necessary to fall back on an ordinal-conversion strategy and to use this strategy with the adopted metric framework. (Altering the framework once it is established is problematic because the numerical values assigned to the items under different metric schemes would often fail to reflect the underlying ordering relations thought to hold between target items.) Here the participant might decide for any number of reasons (see below) that Germany has a large population. The conversion component would then select a value from that portion of the range that has been assigned to large population countries—say, 150 million.

Germany does have a large population—it is the 12th largest country in the world—but with an actual population of 83 million, an estimation of 150 million is far from accurate. In this case, an inaccurate estimate was produced because the range was partitioned in the wrong way. Mistaken ordinal judgments can also yield inaccurate responses. For example, a person who believes that Nigeria has a small population and that such countries have fewer than 20 million people would produce an inaccurate estimate as well. But in this case, it was the mapping decision that was wrong—Nigeria, with a population of 127 million people, has a large population. And of course it is also possible to be right for the wrong reasons; for example, a person who believes that the Netherlands has a medium-sized population might provide an accurate estimate of its population (e.g., 16 million)

if he or she partitioned the range in a very conservative manner (e.g., assigned values between 10 and 20 million to mid-sized countries).

a. Setting and Measuring the Metric The ordinal-conversion strategy appears to involve a preparatory stage during which individuals draw on relevant metric beliefs to define and partition a response range. At this point, the processes required to set and partition the range have not been investigated systemically. However, the work on seeding has demonstrated that numerical facts can be used to improve the accuracy of metric beliefs (see below). We have argued that this effect comes about because people are capable of inducing the metric properties of a target dimension from a set of numerical facts. This is consistent with the research on intuitive averaging; these studies have demonstrated that people are remarkably good at estimating the statistical properties of a set of numbers, even when the numbers are presented at a very rapid rate (Malmi & Samson, 1983; Spencer, 1961). It has been assumed that the same numerical induction process that people employ when they perform a numerical averaging task is used to create and revise metric beliefs.

Given the efficiency of the numerical induction process, one might expect that metric beliefs would be fairly accurate. However, this would also require that people occasionally encounter numerical facts that reflect the metric of a target dimension and that they attend to these facts when they do. The data suggest that this is rarely the case—in general, it appears that people do have a (very) rough sense of the magnitude of the target dimension, but that the metric assumptions are often far from accurate.

It is possible to assess the accuracy of people's metric beliefs and to compare accuracy across items and dimensions by computing two measures: *Signed Order of Magnitude Error (SOME)* and absolute *Order of Magnitude Error (OME)* (Brown & Siegler, 1992, 1993; Nickerson, 1980). SOME and OME are computed as follows:

$$\begin{aligned} \text{SOME} &= \log_{10}(\text{Estimated Value}/\text{Actual Value}) \\ \text{OME} &= |\text{SOME}| \end{aligned}$$

OME converts estimation error to a percentage of an order of magnitude, with small OME values indicating accurate metric beliefs, and large ones indicating inaccurate metric beliefs. Similarly, SOME provides a measure of bias in terms of a percentage of an order of magnitude; a negative SOME indicates that the metric has been set too low, and a positive SOME indicates that it has been set too high.

The OME measures are useful for studying real-world estimation, in part because they minimize the effects of outliers. Outliers are a general problem with this type of research because some distributions (e.g., national populations) are highly skewed, and because the response range often spans several orders of magnitude. As a result, a single estimate can differ from others by orders of magnitude and

thus have an undue influence on conventional accuracy measures (e.g., the mean of the estimates, and the Pearson correlations).

OME and SOME are also attractive measures because they enable direct comparison between different tasks. When such comparisons are made, it is apparent that there are enormous between-task differences in metric accuracy. For example, mean OME is considerably smaller when people estimate city-to-city distances (mean OME = .24) than when they estimate national populations (mean OME = .45; Brown & Siegler, 2001), and it is considerably smaller when they estimate national populations than when they estimate the fatality rates (mean OME = .95; Bostrom & Brown, 2001).

The OME values just reported were computed from initial estimates collected over a representative (or exhaustive) set of target items using a standard 3-phase seeding procedure, in several different experiments. In order to gain a better sense of how metric accuracy varies across dimensions, I have conducted several studies that included questions about a wide variety of real-world quantities (Brown, 2001). Results from one of these are presented in Table I. In this experiment, 125 subjects estimated 23 real-world quantities and used a 1 (no confidence) to 5 (very confident) rating scale to indicate how confident they were in the accuracy of each of their estimates.

The data presented in this table make several points. First, they provide additional evidence for several claims made above. Specifically, it is obvious that metric assumptions are often inaccurate, that accuracy varies greatly from one dimension to the next, and that people are rarely in the position to respond by retrieving prestored numerical facts. Across the 23 items, mean OME ranged from .13 to 2.83, with a median of .68, and the percentage of accurate responses (i.e., responses $\pm 5\%$ of the actual value) ranged from 23 to 0%, with a median of 1%. Second, these data indicate a correspondence between the magnitude of the to-be-estimated values and the estimates themselves. In this study, the rank order correlation between actual value and the median of the estimated values was .81. In other words, although people's metric assumptions may be off by a large amount, it appears that they still tended to respond with relatively small values when the dimension was defined by small magnitudes and with relatively large values when the dimension was defined by large magnitudes. Third, there was a general bias to underestimate the metric; the median SOME for this set of items was $-.53$.

Fourth, it is clear that the magnitude of the to-be-estimated value is not the only factor that determines metric accuracy. This can be seen by comparing mean OME across pairs of items with similar values (e.g., number of goals scored by Gordie Howe versus the average temperature of Venus; the price of an SUV versus the number of square feet in an acre). These data and others (e.g., Jacowitz & Kahneman, 1995) suggest that metric beliefs tend to be accurate when the dimension is familiar and the numerical instances are encountered frequently (e.g., prices, salaries, city temperatures, ages). Finally, it appears that people can

TABLE I
 TWENTY ESTIMATION QUESTIONS WITH ANSWERS, MEDIAN ESTIMATES, MEAN OMEs, SOMEs, HIT RATES,
 AND CONFIDENCE RATINGS

	Mean					
	Actual value	Median estimate	OME	SOME	%Hits	Confidence
How many Olympic medals were won by South Korea in 1996?	27	7	.61	-.57	.01	1.44
How many bicycle patrol officers does the Vancouver Police have?	60	150	.65	.45	.02	1.39
How many episodes of Star Trek: The Next Generation were made?	178	150	.44	-.09	.00	1.59
How many countries belong to the UN?	189	60	.56	-.53	.03	2.02
How many bones are in the human body?	206	208	.31	.09	.23	2.38
What is the average annual snow fall in Prince Edward Island (cm)?	337	40	.93	-.85	.00	1.54
How many goals did Gordie Howe score during his NHL career?	801	600	.40	-.27	.14	1.92
What is the average temperature of the planet Venus (°F)?	870	125	.81	-.69	.02	1.31
How many people died on the Titanic?	1503	1000	.44	-.21	.07	2.03
How many teeth does a great white shark have?	3000	70	1.60	-1.60	.00	1.59
How high is Mt. Everest (meters)?	8850	6000	.68	.00	.06	1.66
How many wild rhinos are there?	11,000	800	1.22	-1.00	.00	1.37
How far is it between Edmonton and Sydney, Australia (km)?	13,299	26,000	.85	.56	.01	1.54
How many islands are in the Indonesian Archipelago?	13,600	15	2.85	-2.85	.00	1.34
How many seats are there in Edmonton's SkyReach Center?	17,100	17,100	.32	-.02	.19	2.67
What is the price of a 2001 GMC Yukon Explorer SLE 4X4 SUV?	41,480	40,000	.13	-.01	.16	2.61
How many square feet are there in an acre?	43,560	400	2.12	-2.03	.00	1.58
How many Canadians died of cancer in 1997?	59,775	23,950	.90	-.35	.01	1.84
What is the land area of Afghanistan (square kilometers)?	652,200	10,000	1.79	-1.72	.01	1.26
What is the current population of Edmonton, Alberta?	668,553	790,000	.45	-.14	.10	2.72
How much is the Japanese Prime Minister's annual salary (US dollars)?	676,000	175,000	.67	-.56	.01	1.54
How many copies were sold of Sarah McLachlan's album "Surfacing"?	7,000,000	1,000,000	1.00	-.94	.00	1.81
How many new cars were sold in the USA in 1999?	8,698,284	1,000,000	1.24	-1.05	.00	1.56
		800	.68	-.53	.01	1.59
		134,421	.91	-.63	.05	1.77

Data drawn from Brown (2001).

judge with some accuracy whether their estimates are “in the ball park” or not. This ability is reflected by ratings which indicate that estimates were generally offered with little confidence and by the strong relation between OME means and the rating means; the rank order correlation between these two measures was $-.67$.

In brief, there is good evidence that metric knowledge varies widely from one dimension to the next and that metric beliefs are generally not very accurate. However, people can distinguish between questions that can be answered with small numbers and those that require very large numbers. This classification of dimensions into small and large can be seen as metric knowledge at its most primitive. In addition, an examination of the well-calibrated dimensions suggests that metric accuracy is related to experience with the dimension and exposure to numerical examples. Of course, these empirical generalizations do not directly address processing issues. At this point, we simply do not know how domain knowledge, numerical beliefs, and response biases are combined to establish a set of provisional range assumptions. Clearly, this is an important question, and one that will require additional research.

b. The Ordinal Component The ordinal-conversion strategy engages two components, one that determines the relative magnitude of the target item and a second that selects a number from the corresponding portion of the response range. The term “component” is used here rather than the term “process” because a number of cognitive processes can be used alone or in combination to produce an ordinal judgment. In this section, I describe five of these. This set includes: ordinal retrieval, ordinal reconstruction, categorical inheritance, guessing, and memory assessment.

i. Retrieval and reconstruction. Ordinal retrieval, ordinal reconstruction, categorical inheritance, and guessing all begin with a search for memory for task-relevant information. Ordinal retrieval serves as the basis for an ordinal judgment when this search encounters a fact that specifies the ordinal value of the target item. For example, a person might simply know that Brazil has a large population. Once this fact is recalled, it is possible to terminate the search and proceed to a numerical conversion process.

Ordinal reconstruction is carried out in much the same way as numerical reconstruction. In both cases, people engage the retrieval–inference cycle, recalling facts related to the target item and drawing plausible inferences from them. What differentiates the two is the presence or absence of numerical facts. When such facts are scarce or nonexistent, the retrieval–inference cycle can deliver little more than a well-grounded ordinal judgment. Nonetheless, informal protocols suggest that this process can be quite lengthy, with participants considering a wide variety of potentially informative facts. For example, a person estimating the population of Germany may know that it has an extremely strong economy and that it played an important role in the history of the 20th century and conclude from these facts that Germany must have a large population.

ii. *Categorical inheritance.* The study of subjective geography has revealed the existence of a special sort of ordinal process, one that falls between ordinal retrieval and ordinal reconstruction (Friedman & Brown, 2000a,b; Friedman et al., in press-b; Kerkman et al., 2001). In these experiments, people are presented with a set of cities and are required to estimate the latitude or longitude of each. Figure 2 presents a *subjective location profile* constructed from data collected in one such study (Friedman & Brown, 2000b, Experiment 1). To construct this figure, test cities were ranked by their mean estimated latitudes and listed from left to right along the abscissa according to their rank. The black circle, the grey circle, and small marker plotted above a city name denote, respectively, the mean of the preseeding estimates elicited by that city, the mean of the postseeding estimates (see below), and its actual population. In this experiment and many others, people's latitudes estimates produce a step function in which cities from a given region are grouped together, but do not overlap with estimates for cities from neighboring regions.

Friedman and I have argued that this pattern indicates that participants primarily rely on regional or categorical knowledge and *categorical inferences* to generate their estimates. We have also identified two sorts of biases in this task: one that leads people to overestimate or underestimate the locations of entire regions (e.g., Mexico), and a second that leads people to compress the range of estimates within a region. The former is attributed to an erroneous partitioning of the range, and the latter to reliance on categorical knowledge in the absence of accurate knowledge of the relative position of items subsumed by the category (Huttenlocher, Hedges, & Duncan, 1991).

What is special about geography is that people divide the world into regions; that they know the ordinal relations that obtained between these regions and correctly believe that regional membership and location are closely linked; and that they have ready access to the regional membership of cities and countries. Under these conditions, the metric can be set by assigning upper and lower bounds to regions in a way that preserves their ordinal relations. Once the metric is established, an estimate can be generated by retrieving the test city's superordinate region and selecting a value that lies within its range. In other words, the target item *inherits* the ordinal properties for its superordinate category, and also its metric properties when they have been defined. Thus, in geography tasks, determining the region or category that a city belongs to is tantamount to assigning it an ordinal value. This process resembles ordinal retrieval because an ordinal decision can be based on a single retrieved fact, and it resembles ordinal reconstruction because the ordinal value is implicit in the retrieved information rather than explicit.

When people have little knowledge of the internal structure of the category (i.e., the relative ordering of items *within* the category), they either pick the mid-point of the region or they spread their responses more or less randomly across it. In contrast, when people know something of the relative ordering of the items

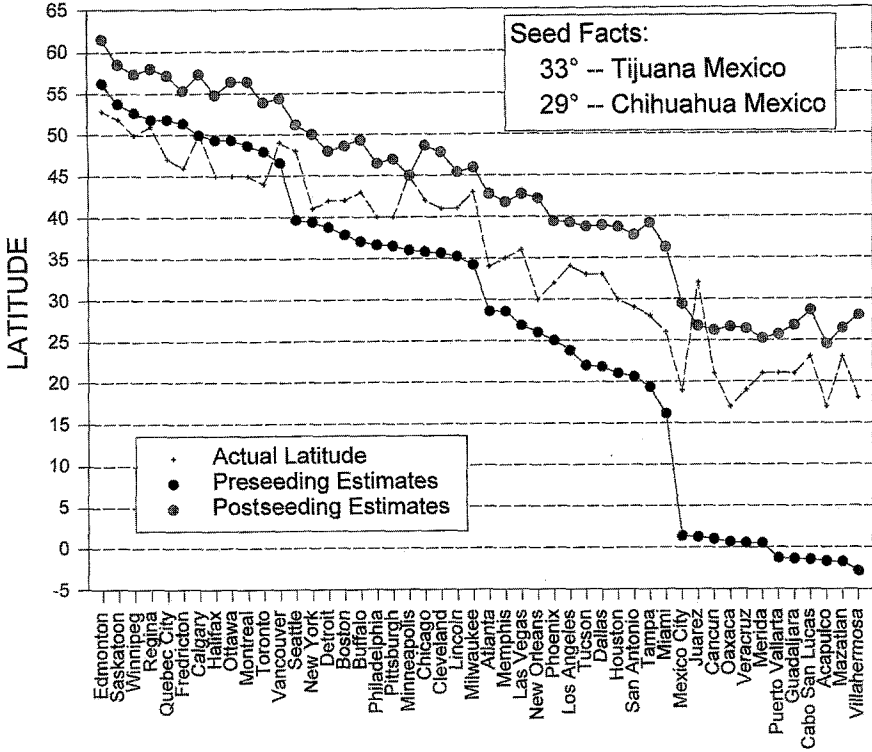


Fig. 2. Subjective latitude location profile for North American cities. For each city, the mean preseeding estimate is represented by the black marker, the mean postseeding estimate is represented by the gray marker, and the actual location is represented by the cross. Postseeding estimates collected following exposure to two Mexican seeds (Tijuana, 33°; Chihuahua, 29°). Data drawn from Friedman and Brown (2001, Experiment 1).

within a category, they display this knowledge by selecting a value from the appropriate part of the subrange associated with the region. The former mode produces a “flat step” and a weak within-region correlation between the estimated and the actual latitudes of the target cities, and the latter produces a “sloped step” and reasonable correlation between estimated and actual latitude. In Fig. 2, estimates for the Mexican cities provide a nice illustration of the former pattern and the estimations for the Canadian cities provide a nice illustration of the latter.

Although latitude estimation has some unique features, other tasks have produced step-like response profiles which indicate reliance on categorical inferences. For example, in a recent experiment, Canadian university students were asked to estimate the current sales prices for automobiles produced by prestige

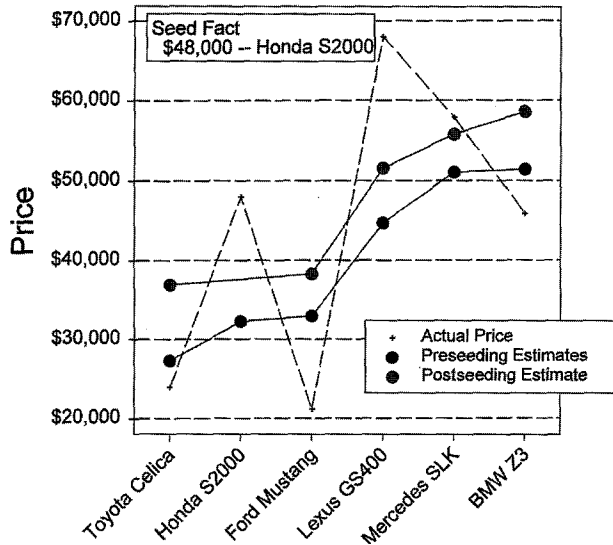


Fig. 3. Subjective price profile for sports cars. For each brand, the mean preseeding estimate is represented by the black marker, the mean postseeding estimate is represented by the gray marker, and the actual price is represented by the cross. Postseeding estimates collected following exposure to the price of Honda S2000 (\$48,000). All prices in Canadian dollars. Data drawn from Murray and Brown (2001).

manufacturers (e.g., BMW, Mercedes Benz) and by nonprestigious manufacturers (e.g., Honda, Ford). The subjective price profile for the sports cars is presented in Fig. 3. This profile indicates that the brands were divided into two price categories and that participants relied on their categorical knowledge to generate their estimates, assigning higher prices to automobiles produced by the prestigious manufacturers and lower prices to those produced by the nonprestigious manufacturers.

Although people were good at estimating the prices of these items, the reliance on categorical knowledge did systemically bias some responses. This point is made very clearly by comparing estimates elicited by the Honda S200 (actual price \$48,000) to those elicited by the BMW Z3 (actual price \$45,900). Although the former was slightly more expensive than the latter, participants apparently treated the S2000 like a typical member of the nonprestige category and the Z3 like a typical member of the prestige category. As a result, the estimated price for the Z3 (\$51,400) was almost \$20,000 greater than the estimated price for the S2000 (\$32,200).

iii. Guessing. In real-world estimation tasks, people occasionally encounter unfamiliar test items. When this happens, they guess. Guesses can be understood by

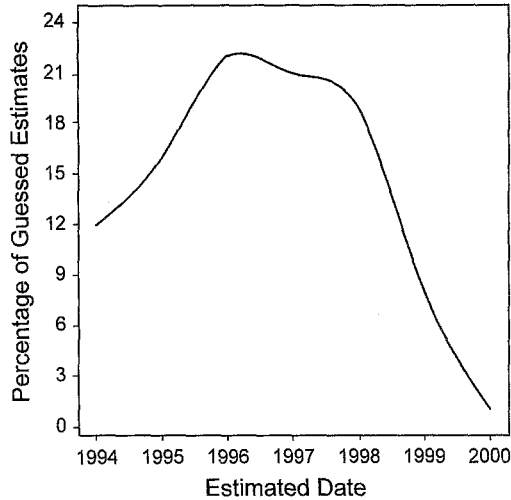


Fig. 4. Guessing function obtained from an event dating experiment. The function represents the percentage of 12,312 guessed responses (i.e., date estimates elicited by unrecognized event descriptions) assigned to each year in the response range. (From Lee, 2001.)

examining responses produced when people indicate that they have no knowledge of the test item, and when this is done, it is clear that it is a systematic aspect to guessing. For example, Fig. 4 presents a guessing function obtained in an event dating experiment (Lee, 2001). In this experiment, participants first rated their knowledge of 67 public events. They were then informed that all events took place between January 1994 and the present and were required to estimate when each target event took place (September, 2000). The function plotted in Fig. 4, which is based on almost 1,700 guessed responses (i.e., estimates elicited by unrecognized events), represents the percentage of guesses that were assigned to each year in the response range. These data clearly indicate that guessing was nonrandom. Rather, it appears that when people guessed, they assumed that the unrecognized events did not happen in the recent past and they had a preference for selecting values near the center of the response range.

Figure 5 presents another guessing function. In this case, the data come from an experiment in which 60 Canadian undergraduates rated their knowledge of 114 countries and then estimated the current population of each (Brown, Cui, & Gordon, in press). The distribution plotted in this figure, which was based on about 1,500 guesses, indicates that there is also a systematic aspect when people estimate national populations, but that guessing in this task differs from guessing observed in dating experiments. Here, it appears that people assumed that obscure countries have small populations and that they acted on this belief by assigning small values

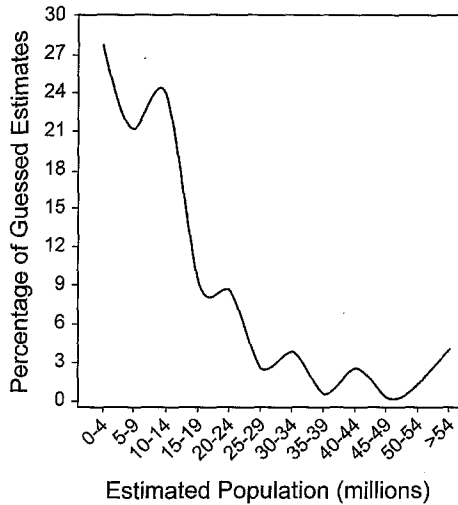


Fig. 5. Guessing function obtained from a population estimation experiment. The function represents the percentage of 12,312 guessed responses (i.e., population estimates elicited by unrecognized country names) falling into each of the specified bins. Data drawn from Brown, Cui, and Gordon (in press, Experiment 1).

to unfamiliar countries. As it turns out, this assumption is a reasonable one, and as a consequence, guessed estimates tended to be fairly accurate; in this experiment, the median of the guessed responses was 10 million, and the median population of countries that were being estimated was 8.7 million.

The point here is that there is a nonrandom aspect to guessing and that guessed responses can under some conditions be quite accurate. The shape of the guessing functions do differ from task to task. This fact rules out the possibility that people rely on some sort of content-free numerical strategy (e.g., pick the middle of the range; pick a small number; distribute guesses evenly across the range) when they fail to recognize the target item. Rather, these differences reflect task-specific and quite reasonable beliefs about the relation between (a lack of) item knowledge and the ordinal value of a target item. For dates, the guessing function suggests that people recognize that events can be forgotten over time and that they assume that an obscure event is one that has been forgotten, and hence one that is not recent (Brown, Rips, & Shevell, 1985). For populations, it may be that guessed and nonguessed estimates are produced by the same process. In the next section I discuss the possibility that people use familiarity or availability as an index of national population.

iv. Memory assessment. Along with the numerical strategies discussed above, ordinal retrieval, ordinal reconstruction, and categorical inheritance depend on the

recovery and evaluation of relevant domain-specific knowledge. Furthermore, it seems that at least some guessing strategies may be informed by an understanding of the relation item knowledge and item value. In addition to these overt processes, there is evidence that implicit factors sometimes play a role in real-world estimation. In particular, an *availability bias* has been observed in several estimation tasks. This bias is defined as a tendency for people to provide larger estimates for well-known items than for less well-known items that have the same actual value. Similarly, it is common to find that estimates elicited by target items correlate more strongly with rated knowledge (a common measure of availability) than with their actual values. Availability biases have been observed when people estimate national populations (Brown et al., in press; Brown & Siegler, 1992, 1993), dates for public events (Brown et al., 1985; Kemp & Burt, 1998; Friedman, 1996), university tuitions (Lawson & Bhagat, in press), and annual revenues of Fortune 500 companies (Brown & Brown, 1987).

The availability bias has been explained by assuming that people are sensitive to the ease with which mental operations such as these are carried out and/or can readily assess how much they know about a given target, and that this evaluation of fluency, familiarity, or associative density can serve to index the relative size of the to-be-estimated quantity. For example, people may believe that better-known countries generally have larger populations than less well-known countries. If so, they may use some process-based measure to gauge how well known the target item is and use the products of this memory assessment process as the basis for an ordinal judgment. Item knowledge and the magnitude of to-be-estimated value are sometimes related (e.g., the rank order correlation between rated country knowledge and actual population is about .35). As a result, these familiarity-based intuitions can serve to inform ordinal judgments. However, because many factors (e.g., recency of exposure, biased news coverage, personal interest, etc.) affect how much people know about a topic and how accessible that knowledge is, an availability bias should be observed when a memory assessment process provides the basis for these judgments (Tversky & Kahneman, 1973).

There are alternative explanations for the availability bias. Suppose that people typically use overt ordinal judgment strategies when they recognize a target item, that they fall back on a guessing strategy when they do not, and that these guessing strategies incorporate reasonable assumptions about what it means for an item to be unknown. For example, a person might infer that unrecognized countries have small populations, that unrecognized hazards are rare, or that unrecognized public events did not happen recently. If guesses are more common when people respond to less well-known items than when they respond to better-known items, and guesses are systemically biased and nonguessed responses are not, then the former should on average produce smaller or less recent estimates than the latter. In other words, systemic guessing could produce an availability bias in the absence of familiarity-based intuition.

There is evidence that suggests that sophisticated guessing is not solely responsible for the availability bias. This point was demonstrated in a recent priming experiment (Brown et al., in press, Experiment 2). In this experiment, test countries were divided into two similar sets. At the outset of the session, participants in one group rated their knowledge of countries in one set and participants in a second group rated their knowledge of countries in the other set. Then, participants in both groups estimated the populations of all test countries.

Assuming that familiarity is reflected in population estimates and that the exposure provided by the initial knowledge rating task would increase familiarity for rated (i.e., primed) countries but not for unrated (i.e., unprimed) countries, it follows that participants should provide larger population estimates for primed countries than for unprimed countries. Thus, the average estimated population for a given test country should be greater when that country had appeared in the initial knowledge rating task than when it had not. This prediction was supported by the data; well over half the test countries displayed the predicted priming effect (also see Brown et al., 1985, Experiments 2 and 3).

Although this result does not rule out sophisticated guessing as a possible source of the availability bias, it does demonstrate that real-world estimates can be primed like other types of nonmnemonic judgments (e.g., fame judgments, Jacoby, Kelley, Brown, & Jasechko, 1989; truth judgments, Begg, Anas, & Farinacci, 1992). This claim, in turn, raises a set of issues concerning the integration of implicit and explicit sources of evidence when both can serve as input to the judgment process. Siegler and I have argued that when domain-specific knowledge and familiarity-based intuitions are available, the former is not necessarily given more weight than the latter (Brown & Siegler, 1992, 1993). Instead, estimates are hypothesized to reflect a weighted blend of competing sources of information, with the weighting of each source determined by its predictive strength (Brunswik, 1955). On this view, familiarity-based intuitions should play an especially important role in determining the ordinal value of a target item, when domain knowledge is sparse, and when these intuitions are more predictive than domain-specific knowledge.

III. Seeding the Knowledge-Base

At its simplest, seeding the knowledge-base requires that a person be exposed to at least one numerical fact and then estimate the value of at least one *transfer* (i.e., nonseed) item. The effects of seeding are assessed by comparing preseeding (or unseeded) estimates to the postseeding estimates. In the typical seeding experiment, postseeding estimates are more accurate than preseeding estimates. However, it is also common to find that the improvement is restricted to measures of metric accuracy and that seeding has little if any effect on the subjective ordering of test items. In other words, seeding almost always leads to a reduction in OME,

but rarely affects the rank-order correlation between the estimated values and the actual values (Brown & Siegler, 1993, 1996, 2001; Friedman & Brown, 2000a,b; Friedman et al., in press; LaVoie et al., in press).

The dissociative effect of seeding on OMEs and correlations has been taken as evidence for the independence of metric knowledge and mapping knowledge (Brown & Siegler, 1993, 1996). Because the numerical-retrieval mode conflates these two types of knowledge, it follows that seeding effects may be limited to estimation tasks that rely on ordinal-conversion processes. And, the differential effect of seeds on measures that reflect metric knowledge and those that reflect mapping knowledge implies that seeding affects the former, but not the latter.

A. THE FEEDBACK/INDUCTION POSITION VS. THE ANCHORING POSITION

This conclusion still leaves open the question of process. How does seeding affect metric beliefs? One possibility is that seeds serve as reference points or anchors, in effect allowing people to shift from an ordinal-conversion mode to a numerical-retrieval mode. Because seed facts are expressed in the correct metric, the reconstructive numerical inferences sanctioned by these facts would also carry valid metric information with them. This is referred to as the *anchoring* position. A second possibility is that seeds provide feedback on the accuracy of preexisting metric beliefs and data necessary for inducing more accurate beliefs. This is referred to as the *feedback/induction* position.

An example is helpful for understanding the anchoring position. Consider a person who initially believed that 20 million people lived in Pakistan and who then learned that there are 129 million Bangladeshis. If this person were required to estimate Pakistan's population again, he or she might recall the population of Bangladesh, recognize that the two countries are similar in relevant respects, infer that Pakistan and Bangladesh have similar populations, and conclude that some 130 million people live in Pakistan. This second estimate is far more accurate than the first—the correct answer is 142 million.

More generally, the anchoring position assumes the following: (a) Seeds are stored in memory during a learning phase. (b) At least one seed is retrieved when a participant provides a postseeding estimate for a transfer item. (c) Similarity between the seed and the transfer item determines whether a given seed is retrieved. (d) Estimates for the transfer item are “drawn toward” or assimilated to the retrieved seed(s). Consistent with Assumption a, participants in seeding experiments typically learn the seeds they have studied (Brown & Siegler, 1993, 1996; LaVoie et al., in press). Assumptions b and c describe how the retrieval–inference cycle and numerical reconstruction are linked. Assumption d restates the core finding in the anchoring-and-adjustment literature; extensive research conducted on this topic has demonstrated that transfer values are almost always assimilated to

anchor values (e.g., Jacowitz & Kahneman, 1995; Strack & Mussweiler, 1997; Tversky & Kahneman, 1974).

Taking advantage of the terminology introduced above, anchor position implies a mode switch; on this view, when seeding effects are observed, participants who relied on ordinal-conversion processes to generate their preseeding estimates use numerical reconstruction to generate their postseeding estimates. In contrast, the feedback/induction position holds that ordinal-conversion mode is used to produce both pre- and postseeding estimates and that seeding effects come about because the people have used the seeds to correct the metric assumptions that define the response range.

More concretely, consider a person who initially believed that most countries have populations around 10 million and who then learns that the actual population of a representative set of countries is around 30 million. In this situation, this participant is likely to realize that the upper bound of the subjective response range has been too small or/and that range has been partitioned too conservatively. In response to this feedback, it would make sense to revise the range assumptions in a way that brings them in line with the metric information conveyed by the seed facts. This revision of the range should have two consequences; first, the postseeding estimates should be more accurate than the preseeding estimates, and second, the participant should come away with an improved set of metric beliefs.

Four lines of evidence now support the claim that a feedback/induction mechanism produces seeding effects. First, seeding effects persist even after seed facts are forgotten (Brown & Siegler, 1996; LaVoie et al., in press). This was demonstrated in an experiment designed to examine the long-term impact of the seeding. In this experiment, participants first estimated the populations of 99 countries; next, they learned the actual populations of 24 of the countries; then they estimated the populations of all 99 countries a second time. Four months later, the same participants returned to the laboratory and produced a third set of estimates for the 99 countries. For seed countries, mean OME prior to the learn phase was .56; it dropped to .08 immediately following the learning phase, and increased to .34 over the 4-month delay. Comparable values for the transfer countries were .65, .41, and .42. In other words, although the seed facts tended to be forgotten over time, the transfer estimates obtained 4 months after seed facts were learned were no less accurate than estimates produced immediately after the seeding procedure, and both were far more accurate than pretest estimates.

The anchoring position holds that the accuracy of estimates should be directly linked to knowledge of the reference points. Thus, this view incorrectly predicts that transfer estimates collected in the 4-month condition should have been less accurate than those obtained immediately after seeding. In contrast, the feedback/induction position holds that knowledge of the actual values of individual seed facts is unnecessary once the metric beliefs have been updated. Thus, the feedback/induction position correctly predicted that delayed postseeding estimates would remain more accurate than the preseeding estimates, even after the seed facts had been forgotten.

A second line of evidence in favor of the feedback/induction position comes from research that has demonstrated that exposure to the numerical values alone can produce effects that are as robust as those that are obtained when people learn a corresponding set of seed facts. Recently two separate studies have shown this effect, one concerning population estimates (LaVoie et al., in press) and the other fatality rate estimates (Bostrom & Brown, 2001). The LaVoie study involved several between-subjects conditions. These included a condition in which participants were presented with 24 seed facts, one in which they were given the numerical values associated with the mean and range of the seed countries, and one in which they received only the mean value of the set. All three groups benefited from the information they received and displayed large reductions in OME. What was surprising was that the postseeding OME for the seed-24 group (.38) was only slightly smaller than the postseeding OME for the mean and range group (.41) and the mean-only group (.45).

A similar result was obtained in a study in which participants were presented with a set of risks and required to estimate the relative frequency of each. After producing an initial set of estimates, participants in one group, the seed group, were given the actual relative frequencies of a subset of the hazards, and participants in a second group (the risk-scale group) were given a "community health scale." The community scale associated levels of risk with appropriately sized social units for each of the 7 orders of magnitude, ranging from 1 in 100 ("1 in 100 means that you can expect 1 fatality from this hazard on a street") to 1 in 100 hundred million ("1 in a hundred million means that you can expect 1 fatality from this hazard in a large country"). The seeds facts were selected to correspond to the seven community risk statements. For example, participants in the seed group were given the information that the annual mortality rate associated with all causes of death was 1 in 100 and that the annual mortality rate for whooping cough was 1 in 100 million.

After receiving this information, participants in both groups provided a second set of estimates. As in the LaVoie et al. (in press) study, posttest estimates were much more accurate than the pretest estimates, and the format manipulation had a negligible effect on performance. For both groups, the mean pretest OME was greater than 2.0, and mean posttest OME was about 1.2.

On the anchoring view, when people have learned a set of seed facts, they shift from an ordinal-conversion mode to a numerical-retrieval mode and come to rely heavily on numerical reconstruction. This process requires not only that a numerical value is retrieved, but that the retrieved value has a known or inferable relation to the target item. On the one hand, the similarity of categorical relations that obtain between the seed items and the target items should enable the retrieval and interpretation of recently encountered numerical values. On the other hand, the absence of seed items should make it difficult if not impossible for people to relate the target items to numerical values when they are detached from specific categories or instances. Thus, the anchoring position incorrectly predicts that people can take advantage of numerical values only when they are presented as seed facts and fails

to account for the performance enhancing effects of the range and mean values in the LaVoie study and of the community risk scale in the Bostrom and Brown study.

The feedback/induction position holds that seeding effects come about because the seed facts enable people to assess and when necessary to revise their metric knowledge. More generally, this view assumes any information that can be used to improve the accuracy of mistaken metric beliefs should improve estimation performance and that the format of the feedback is unimportant. Pretest results provided clear evidence that participants in both the LaVoie et al. (in press) and the Bostrom and Brown (2001) studies were operating with inaccurate range assumptions and a poor understanding of the relevant magnitudes. Under these conditions, the feedback/induction position correctly predicted that these individuals would benefit from exposure to numerical values that correctly defined the boundaries of the response range and/or its central tendency.

B. TWO CLASSES OF SEEDING EFFECTS

1. *Category-Based Seeding Effects*

The central finding in the anchoring literature is that postcomparison estimates assimilate to the comparison value. By extension, the anchor position predicts that that transfer values should only assimilate to seed values. This prediction has been disconfirmed by two lines of evidence. In this section I review the evidence that seeds produce robust contrast effects when they are used to repartition the response range. I also speculate on the conditions required for seeds to produce a repartitioning of the range and sketch the logic that has allowed us to predict when and how seeds of this type will affect performance. The following section focuses on a more general form of seeding, one that is observed when people use seed information for revision, rather than repartition, of their response range. In this section, I present evidence demonstrating that this type of seeding also produces contrast effects.

Both the geography project and pricing project demonstrate that postseeding estimates for some target items are drawn toward the seed values while others shift away from them. This predictable mixture of assimilation and contrast is apparent in Figs. 3 and 4. The postseeding estimates presented in Fig. 3 (represented by the gray markers) were obtained from participants who had learned the actual latitudes of Tijuana (33°) and Chihuahua (29°), and the postseeding estimates presented in Fig. 4 were obtained from participants who learned the actual price of a Honda S2000 (\$48,000). In both experiments, items from the seeded category (Mexico and nonprestige brands) moved in the direction of the seed facts, but the items from the adjacent category shifted away from them (southern United States and prestige brands).

The feedback/induction position, but not the anchoring position, can readily account for the contrast effects observed in these and other experiments. On this

view, the seeding effects apparent in these figures occurred because the seeds caused people first to redefine the range associated with seeded category and then to adjust the range associated with neighboring categories in order to keep the values assigned to items from these categories from overlapping.

It is possible that this type of contingent repartitioning of the range will be observed only under the following conditions.

1. Target items can be readily and accurately identified as members of particular categories.
2. People have fairly uniform sets of beliefs about the response range either because the target dimension is familiar (e.g., automobile prices) or because it has been explicitly defined for the participants (e.g., latitude and longitudes).
3. Categories have a known ordering and are believed to cover mutually exclusive portions of the response range.
4. When the range is initially partitioned, numerical values are assigned indicating the upper and lower bound of categories and/or their central values.
5. Categorical (inheritance-based) inferences play a primary role in the task. In other words, when people generate numerical estimates, they first determine the item's category membership and then select a value that falls within the range of values assigned to the relevant category.

Under these conditions, seed values provide feedback on the accuracy of the range of values assigned to the seed's category and also information about the direction and size of the adjustment necessary to produce a more accurate partitioning of the range. When the seed values are greater than the initial values estimated for the seed items (or for other comparable items from the seeded category), people typically respond by shifting the subrange up the scale. The opposite happens when the seed values are smaller than the initial values assigned to the seed items. These shifts have direct and obvious consequences for all items from the seeded regions. When category boundaries for the seeded region are shifted in one direction, the postseeding estimates for items from that category are also shifted in that direction.

In addition, redefining the range associated with the seeded category can and often does trigger predictable revisions in the ranges associated with unseeded categories. There are several possibilities here. Seeds should have their most far reaching effects when the range is divided in a way that is both mutually exclusive and *exhaustive*. Here, a change in the range associated with one category would necessitate changes in the ranges associated with all other categories.

A second more common situation occurs when the seeded category is considered to be *strictly adjacent* to a neighboring category (i.e., when the upper bound of one category is considered identical to the lower bound of the adjacent category). For example, North Americans certainly know that the United States and Mexico are part of the same land mass and believe that the southern United States is north of Mexico. They also tended to locate Mexico and the southern United States too far

to the south (Friedman & Brown, 2000a,b; Friedman et al., in press-b). As noted above, when people learned the actual latitudes of Tijuana and Chihuahua, they shifted cities from both the seeded (Mexico) and the unseeded region (the southern US) to the north (Fig. 3). This is also true when they learn they were seeded with two cities from the southern United States (Orlando FL, 29° and San Diego, CA 33°).

Friedman and I have argued that these effects occurred because participants first used the seed facts to update their beliefs about the range of the seed region. Then, because the upper bound of one region is assumed to be the lower bound of the other (i.e., the regions are strictly adjacent), it was necessary to adjust the range of the neighboring region. This repartitioned range then served to guide the conversion process as participants again transformed their categorical (regional) knowledge to a set of numerical estimates.

The work on geography also indicates that seed effects can propagate to *conceptually coordinated* categories. These are categories that cover the same portion of the response range, though in parallel ordinal structures. To take a concrete example, latitude estimates for the Old World cities and New World cities overlap to a large extent and the estimates for the southern United States and southern Europe tend to be very similar. We have demonstrated that seeding one of these regions produces a comparable shift in the other (Friedman & Brown, 2000b, Experiment 2). This effect occurred because people believe that two regions covered the same range of latitudes. Given this belief, a seed-induced revision to the range associated with one region implies that the range associated with the other regions should also be revised.

We have also argued that two principles, *coherence* and *inertia*, determine whether seeding effects will propagate from a seeded category to other categories, in the absence of strict adjacency. The coherence principle holds that people will adjust the metric values associated with an unseeded category when this adjustment is necessary to maintain the ordinal relations assumed to obtain between neighboring categories. The principle of inertia is the inverse of the coherence—it states that in the absence of strict adjacency, people will *not* adjust the metric values associated with unseeded categories when the coherence of the representation is not compromised by the modifications to the range of a neighboring category.

A geography experiment focused on the Old World provided a clear demonstration of these principles at work (Friedman & Brown, 2000a, Experiment 2). In this experiment, Canadian undergraduates estimated the latitudes of cities in Europe and Africa. One group then learned the location of two southern European cities (Lisbon, 39° and Athens, 39° north) and another group learned the location of two north African cities (Tunis, 37° and Algiers, 37° north). All participants then provided a second set of estimates for all test cities. Consistent with previous research, southern Europe and Africa formed distinct regions, and the cities in both regions were placed far to the south of their actual locations. Following exposure to the southern European seeds, estimates for cities in southern Europe moved north, but

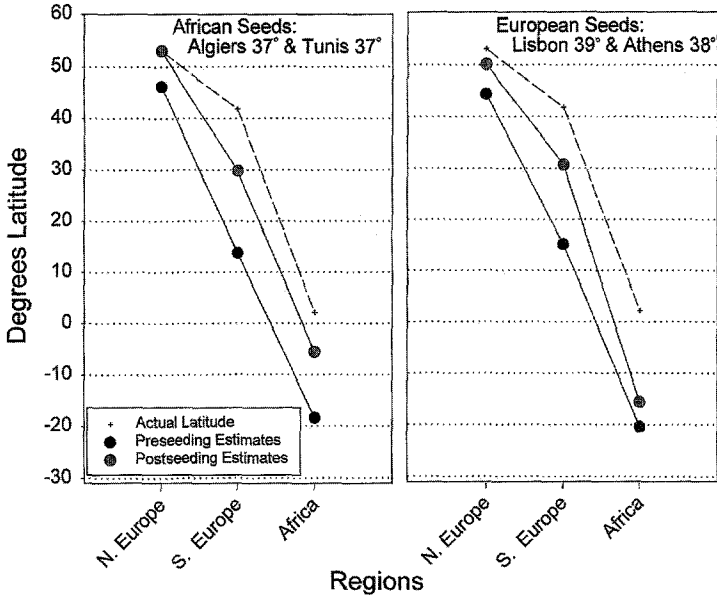


Fig. 6. Region means for Old World cities collected from the African-seed group (left panel) and the European-seed group (right panel). For each region, the preseeding estimate is represented by the black marker, the postseeding estimate is represented by the gray marker, and the actual location is represented by the cross. For the African-seed group, postseeding estimates were collected following exposure to two African seed cities (Tunis, 37° and Algiers, 37°); for the European-seed group, postseeding estimates were collected following exposure to southern European cities (Lisbon, 39° and Athens 39°). Data drawn from Friedman and Brown (2000a, Experiment 2).

the estimates for cities in Africa did not. In contrast, the African seeds affected both Africa and southern Europe in the same way: postseeding estimates for cities from both regions shifted about 15° to the north of the preseeding estimates (see Fig. 6).

An examination of the preseeding estimates provides a starting point for an explanation for this asymmetrical seeding effect. As noted above, people who took part in this experiment treated southern Europe and Africa as separate regions; they knew that southern Europe is north of Africa; and their estimates for cities from both regions were strongly biased to the south. It also seems reasonable to assume that these participants were aware of the existence and relative location of the Mediterranean Sea. Under these conditions, the African seeds had two effects. First, they indicated that the preseeding estimates for the African cities were too far to the south and implied that the range associated with Africa should be shifted to the north. Second, because the values associated with African seeds were north of most preseeding estimates elicited by southern Europe, these seeds also carried implications for the location of southern Europe. The implication here was that

southern Europe would have to be shifted to the north to ensure that the postseeding estimates still reflected the belief that southern Europe is north of Africa.

The revision of the range associated with southern Europe is consistent with the operation of the coherence principle. On this view, participants who were presented with the African seeds updated the range associated with southern Europe because the failure to do so would have contradicted the firmly held beliefs about the position of Africa relative to Europe and not because they believed that the two regions were strictly adjacent. Evidence for this latter point comes from the southern European seed condition. If participants who received the southern European seeds had treated southern Europe and Africa as being strictly adjacent, then a seed-induced shift in the subjective location of the former should have been followed by a comparable shift in the location of the latter. In other words, the estimates for Africa should have been “dragged up” by the southern European seeds in the same way that Mexican estimates were dragged up by southern United States seeds (Friedman & Brown, 2000b, Experiment 1). Instead, the estimates for the African cities were unaffected by the southern European seeds, though the postseeding estimates for the southern European transfer cities were far to the north of the preseeding estimates.

Given that the two regions are separated by a body of water and thus not strictly adjacent, relocating the lower bound of southern European had no clear implications for the location of the upper bound of Africa. In addition, although the southern European seeds indicated that the range assigned to the seeded region needed to be revised, the updated set of range assumptions left the ordinal relations between southern Europe, the Mediterranean, and Africa intact. Thus, neither adjacency nor coherence provided a rationale for participants in the southern European seed group to revise their beliefs about the location of Africa. In the absence of such a rationale, the principle of inertia is in effect; simply put, these people did not update their beliefs about the location of Africa because they did not have to.

It is worth noting that it is possible to use the notions of coherence, adjacency, and inertia to generate additional predictions about the effects of seeds on geographical knowledge. To take one example, we would expect seeding with cities from central rather than northern Africa (e.g., Mogadishu, Somalia, 2° north and Kampala, Uganda, 0°) would lead participants to produce more northerly estimates for transfer cities from Africa, but would have no effect on the location of cities in southern Europe. In this case, the values conveyed by the African seeds would not violate the belief that southern Europe is north of Africa and thus would fail to trigger a coherence-driven revision of the southern Europe range.

In principle it should also be possible to employ these transfer notions to investigate the representation of knowledge in other categorically structured domains. For example, we already know that people classify automobile brands into two categories and that learning about the price of the Honda S2000 increased postseeding estimates for the seeded category (nonprestige brands) and the unseeded category

(luxury brands). This is analogous to the effect of Mexican seeds on the southern United States and African seeds on southern Europe, which means that the transfer could have been mediated by adjacency or coherence. It should be possible to determine whether the two price categories are considered to be strictly adjacent by seeding with a very expensive prestige sports car (e.g., a Lexus GS400, \$68,000) or a very inexpensive nonprestige sports car (e.g., a Ford Mustang, \$21,000). If adjacency holds, the former should increase the prices of cars from both categories and the latter should decrease them. If adjacency does not hold, the seeds should affect only the transfer items from the seeded category.

Finally, it should be pointed out that it is also possible to select seeds in such a way that removes estimation bias. This was done in a recent experiment in which Canadians from Alberta and Americans from Texas estimated the latitudes of Canadian, American, and Mexican cities (Friedman et al., in press-b). They were then presented with the latitudes of Ottawa (45°), Minneapolis (45°), Dallas (33°), and Tijuana (33°) and provided a second set of estimates. As in prior studies, the preseeding estimates for Mexico and the southern United States were biased to the south, with the latter displaying a greater bias than the former. In addition, the Texans initially placed the Canadian cities far to the north of their actual locations, though the Albertans did not. All of these biases were eliminated in the postseeding estimates. It may be too soon to generalize from these findings to a set of prescriptions for selecting an optimal set of seed facts. But these results do suggest that it is important to seed each of the functional categories and that it is a good idea to select seeds that denote category and/or range boundaries.

2. *Revising the Metric*

In the last section, I focused on a set of seeding phenomena that appear to occur only when people have an accurate understanding of the response range and when they primarily rely on categorical (inherence-based) inferences to generate their estimates. Because metric beliefs are often inaccurate and category knowledge typically does not map onto target dimensions in a one-to-one manner, tasks that meet these requirements are uncommon. Nonetheless, seeding effects are readily obtained. The processes that produce most of effects are probably quite simple. As noted above, exposure to the seed facts can provide feedback about the accuracy of the current range of assumptions. When it is clear that these assumptions are inaccurate, the numerical values conveyed by the seeds can be used to induce a more appropriate range and/or a more appropriate partitioning of the range.

A core prediction of these claims is that there should be a strong negative correlation between the magnitude of the initial estimates for the seed items and the magnitude of change produced in the estimates for the transfer items. Specifically, when the actual value of the seed items (S_A) is *greater* than initial estimates for the seed items (S_1), then the postseeding estimates for the transfer items (T_2) should

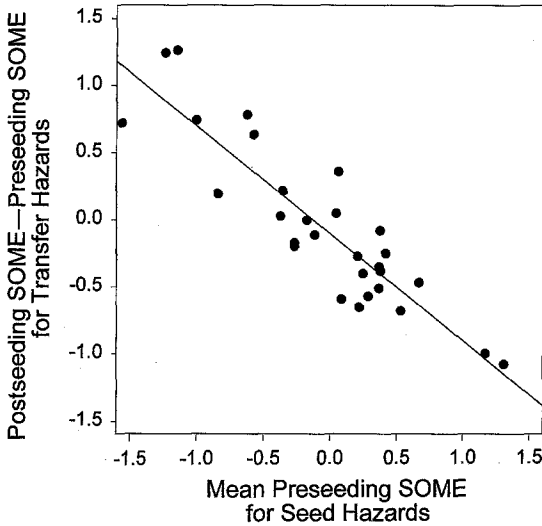


Fig. 7. Mean change in SOME for transfer hazards as a function of the mean SOME for the seed hazards. Data from Bostrom and Brown (2001).

be *greater* than the preseeding estimates for the transfer items (T_1); when S_A is *less* than S_1 , T_2 should be *less* than T_1 ; and the size of the change in the transfer estimates should reflect the degree to which the values for the seed estimates were initially under- or overestimated.

The predicted correlation has been observed when people estimate national populations (Brown & Siegler, 1993, 2001), city-to-city distance (Brown & Siegler, 2001), and fatality rates (Bostrom & Brown, 2001). Data from the latter study are presented in Fig. 7. Each point in this figure represents data from one participant. Each of these participants estimated the number of Canadians who die each year from each of 40 causes; they then learned the actual fatality rates associated with 8 of these hazards and provided a second set of estimates. To construct this figure, three mean SOMEs were computed for each participant, one over the initial seed estimates (S_1), a second over the initial transfer estimates (T_1), and a third over the postseeding transfer estimates (T_2). Then T_1 was subtracted from T_2 providing an indication of the direction and magnitude of the seed-inducing change; thus, a positive value of this difference indicates postseeding estimates were on average larger than the preseeding seeding estimates, and a negative value indicates the opposite.

This difference score is plotted against S_1 in Fig. 7. The correlation between these values was $-.85$, indicating that individuals who initially underestimated fatality rates for the seed hazards increased the magnitude of their postseeding estimates, and those who initially overestimated the seed values decreased the estimates for the transfer items. In addition, although S_1 and T_1 were strongly

correlated ($r = .88$), S_1 and T_2 were not ($r = -.13$). This indicates that people converged on roughly the same metric for their postseeding metric and that they did this regardless of their initial metric assumptions.

For reasons laid out above, this pattern of results is consistent with the feedback/induction position. However, an anchoring account can also explain them. On this view, seeding allows (causes) people to shift from the ordinal-conversion mode to the numerical mode. If so, when people generate their posttest estimates, they would recall at least one seed fact and use it as a numerical anchor or reference point. Given that S_1 and T_1 are highly correlated, this means that people who initially underestimated S_1 and T_1 will be anchoring on relatively large values when they use the seed values to construct their T_2 estimates, and those who initially overestimated S_1 and T_1 will be anchoring on relatively small values. Assuming that T_2 assimilates to the anchoring, it follows that T_2 should be greater than T_1 when T_1 was initially underestimated, and that T_2 should be less than T_1 when T_1 was initially overestimated. This of course is what Fig. 7 shows.

Siegler and I devised a variant of the standard seeding that was capable of distinguishing between these two accounts (Brown & Siegler, 2001). As in most other seeding experiments, we used a 3-phase procedure. However, we selected seed facts whose actual values were less than participants' initial estimates of their values, but were greater than their initial estimates for the transfer items, that is, $T_1 < S_A < S_1$. Under these conditions, the feedback/induction hypothesis predicts that postseeding transfer estimates should shift *away* from the anchors' values, which implies $T_2 < T_1$. This is because the exposure to these particular seed values would imply that the initial estimates tended to be overestimated. If people recognize this fact, they should respond to it by revising their metric assumptions downward. In contrast, the anchoring position predicts that postseeding estimates for transfer items should be drawn *toward* the actual value of the seed items, which implies that $T_2 > T_1$. This prediction is based on the general finding that to-be-estimated values assimilate to available anchor values.

These predictions were tested in two experiments: in one, people estimated national populations and in the second, they estimated city-to-city distances. In both experiments, we are able to select seed items so that a reasonable number of participants produced the "split" pattern that was critical for assessing the competing predictions (i.e., $T_1 < S_A < S_1$). And in both, participants who displayed this critical pattern were much more likely to decrease their estimates than to increase them. Thus, these experiments provide a fourth line of evidence in support of the view that people respond to seed facts by adjusting their metric assumptions and against the view that seeding fosters a more reconstructive approach to real-world estimation.

3. Additional Points

There are three additional issues that are worth mentioning before concluding this section. The first concerns the number of seeds required to obtain a seeding

effect. A fair amount of evidence has accumulated that even a single seed fact can affect the magnitude of people's real-world estimates. Specifically, single seeds, typically presented as examples prior to an estimation task, have been found to influence population estimates (Brown et al., in press, Experiment 1; LaVoie et al., in press), estimates of fatality rates (Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978, Experiment 3), and longitude estimates (Friedman & Brown, 2001a, Experiment 3). Additional demonstrations of the single-seed effect have been reported in experiments in which participants were presented with several different real-world estimation questions and were required to read a seed fact before producing an estimate (Beck & Carlson, 1998; Brown, 2001; Friel & Carlson, 2000). Although there is some evidence that the robustness of the seeding effect increases with the number of seeds presented (LaVoie et al., in press), these single-seed effects are important because they indicate that people are very sensitive to the numerical implications carried by seeds and that they are often so uncertain of the metric that they are willing to revise their assumptions on the basis of a very small amount of information.

The second point deals with the relation between seeds and mapping knowledge. As noted above, exposure to seed facts generally does not affect the relative ordering of transfer items. This fact has been taken as evidence that metric and mapping knowledge are independent. However, there have been two demonstrations that seeds can influence the rank-order correlation between the estimated and actual values. One of these dealt with population estimates (Brown & Siegler, 1993, Experiment 3) and the other with estimates of college tuition (Lawson & Bhagat, in press, Experiment 1). In both studies, seeds were selected from specific categories in a way that emphasised the existence of a relation between category membership and the magnitude of the to-be-estimated values. And in both studies, a followed-up experiment demonstrated that seed facts are not necessary to alter the correlation between estimated and actual values. Rather, changes in mapping accuracy could be achieved by presenting participants with statements that either captured the generalizations implied by the seed set (i.e., "People tend to overestimate the population of European countries and underestimate the population of Asian countries") or facilitated an accurate assignment of a target item to a functional category (e.g., "Private schools tend to be named after people or have religious connotations"). This suggests seeds will affect measures of mapping accuracy when they indicate that categorical knowledge can serve as the primary cue to an item's ordinal value, when they provide the basis for determining the ordinal relations that hold between the functional categories, or when they facilitate categorization of the target items.

The final point concerns the conclusion that seeding effects are produced by a feedback/induction process and not by an anchoring process. Again, this conclusion is based on four types of evidence. Briefly, we have found (a) seeding effects can persist for a long time even when the seed facts have been forgotten; (b) experimenter-provided numbers can affect postseeding estimates in much the

same way that numerically comparable seed facts do; (c) contrast effects are common when seeds are used to repartition the range; and (d) contrast effects can also be observed when seeds are used to reset the metric.

Despite the existence of much evidence favoring the feedback/induction hypothesis, there are empirical, as well as intuitive, grounds to suspect that anchoring may play a role in seeding. Specifically, LaVoie et al. (in press) recently collected written strategy reports from people who had just completed a population seeding experiment. Nine of 12 participants who learned a set of seed facts “reported using the populations given to base their estimates on.” Interestingly, 7 of 12 participants who were presented only the numerical values (in the absence of the country names) indicated the same thing. On the one hand, these data suggest that participants in the seed group may have used the seeds in a (re)constructive manner. On the other hand, for reasons laid out above, it seems unlikely that numerical values stripped of their referents could be used to support numerical reconstruction. This in turn raises the possibility that participants in both groups used the numerical values they were presented with in some other way. For example, these numbers may have been used to adjust their metric assumptions and then to define the boundaries that separated different portions of the range. Or, they may have served as examples of the types of numbers that could pass as a reasonable response. Thus, despite the intuitive appeal of the anchoring position, there is still no compelling reason to believe that seeding causes people switch from ordinal conversion to numerical reconstruction.

IV. Conclusion

In this article I have attempted to lay out a broad framework for understanding real-world estimation and seeding effects. The core contentions of this framework can be summarized as follows:

1. Plausible reasoning and the retrieval–inference cycle play a central but varied role in real-world estimation.
2. There are two basic estimation modes: numerical retrieval and ordinal conversion.
3. People use numerical-retrieval strategies when the target items are typically directly or indirectly associated with at least one potentially relevant numerical fact.
4. When such facts are scarce, people rely on ordinal conversion.
5. Ordinal conversion requires a preparatory stage during which people set the metric (i.e., define and partition the response range).
6. Once the metric is set, estimates are generated by determining the ordinal value of the target item and then selecting a value from the appropriate portion of the range.

7. Many different processes may be used to determine the relative magnitude of the target item, including one that directly equates a target item's category membership with its ordinal value.
8. Exposure to seed facts causes people to revise their metric assumptions and/or repartition the range.
9. These changes are produced by a feedback/induction process rather than an anchoring process.
10. In some situations, estimation profiles and the seeding effects can be used to explore the structure and nature of domain-specific knowledge.

This set of claims captures much of what is known about real-world estimation and seeding. This does not mean that there are not many issues remaining. Some of these concern developing a more detailed understanding of the processes that are engaged when people generate numerical estimates. For example, I have argued that ordinal conversion is the default estimation strategy and that the application of this strategy typically requires people to draw on their rather vague metric beliefs to define and partition the response range. It is clear that a complete description of real-world estimation will be needed to explain how this essential task is accomplished and that additional research will be required to achieve this goal.

Other issues concern various extensions of this approach. My colleagues and I have already begun to conduct research designed to examine the development of the knowledge structures and estimation strategies (Kerkman et al., 2001) and the nature, origins, and implications of cross-cultural differences in estimation performance (Brown et al., 2002; Friedman et al., *in press-b*; Kerkman, Norris, & Stea, 2001). Another way to extend this project is to begin the systematic study of additional estimation tasks. In the past, I have found that each new task has its unique elements and some features in common with other tasks. As a result, understanding a wider range of tasks should contribute additional detail to the current framework, while strengthening the inductive claims that it embodies.

In addition to studying new estimation tasks and considering the developmental and cross-cultural implications of this work, it will also be useful to determine whether the current approach accounts for other types of estimation phenomena (e.g., anchoring effects, boundary effects) and other uses of real-world knowledge (i.e., non-numerical judgment, opinion formation and revision, etc). And it will be necessary to reconcile the current framework with other potentially relevant theories such as Anderson's feature integration theory (Anderson, 1981), Huttenlocher's category model (Huttenlocher et al., 1991), and Gigerenzer's fast and frugal heuristics perspective (Gigerenzer, 1999; also see Brown, 2002; Friedman, 1993; Jacoby & Brooks, 1984; Parducci, 1963; Poulton, 1989). Finally, some clear predictions fall out the framework. For example, seeding should have a minimal effect on estimation performance when people can rely on numerical-retrieval strategies, and there should be little evidence for postseeding mode shift

when seeding effects are observed. Direct tests of such predictions will contribute to an improved understanding of the relation between domain knowledge and strategy selection.

Admittedly, much work remains to broaden and deepen our understanding of real-world estimation and to determine how a theoretical account of this class of behaviors fits with a more general theory of complex cognition. Nonetheless, I would argue that we know much more about these issues now than we did 10 years ago, and that this fact demonstrates that it is not only possible but desirable to use experimental methods to study the organization and utilization of real-world knowledge.

ACKNOWLEDGMENTS

This work has been supported by NSERC of Canada. Much thanks to Bob Siegler and Alinda Friedman who have contributed enormously, and for many years, to every phase of this project. I also thank Fred Conrad, Ann Bostrom, Peter Lee, and Kyle Murray for their essential contributions to our various collaborations. Correspondence should be addressed to Norman R. Brown, Department of Psychology, University of Alberta, Edmonton, AB, T6G 2E9. E-mail: norman.brown@ualberta.ca

REFERENCES

- Anderson, N. H. (1981). *Foundations of information integration theory*. New York: Academic Press.
- Beck, M. R., & Carlson, B. W. (1998, May). *Anchoring effects in numerical estimation with verbal anchors*. Poster presented at the Annual Meeting of the American Psychological Society, Washington, DC.
- Begg, I. M., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: Source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology: General*, *121*, 446–458.
- Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioral frequency questions. *Journal of Consumer Research*, *14*, 280–288.
- Bostrom, A., & Brown, N. R. (2001). *Estimating and revising subjective fatality rates*. Manuscript in preparation.
- Brown, N. R. (1990). Organization of public events in long-term memory. *Journal of Experimental Psychology: General*, *119*, 297–314.
- Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1539–1553.
- Brown, N. R. (1997). Context memory and the selection of frequency estimation strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 898–914.
- Brown, N. R. (2001). *Anchors as seeds: A Gricean perspective on the post-comparison anchoring effect*. Manuscript in preparation.
- Brown, N. R. (2002). Encoding, representing, and estimating event frequencies: A multiple strategy perspective. In P. Sedlmeier & T. Betsch (Eds.), *Frequency processing and cognition*. pp. 37–53. Oxford: Oxford University Press.
- Brown, N. R., & Brown, S. I. (1987). *Estimating sales for Fortune 500 companies*. (Research Report 12475). Yorktown Heights, NY: IBM.
- Brown, N. R., Cui, X., & Gordon, R. (in press). Estimating national populations: Cross-cultural differences and availability effects. *Applied Cognitive Psychology*.

- Brown, N. R., Friedman, A., & Lee, P. J. (2001, November). *Distance estimates reflect objective and subjective geographical categories*. Talk presented at the 41st Annual Meeting of the Psychonomic Society, Orlando, FL.
- Brown, N. R., Rips, L. J., & Shevell, S. K. (1985). Subjective dates of natural events in very long-term memory. *Cognitive Psychology*, *17*, 139–177.
- Brown, N. R., & Schopflocher, D. (1998). Event cueing, event clusters, and the temporal distribution of autobiographical memories. *Applied Cognitive Psychology*, *12*, 305–319.
- 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). Long-term benefits of seeding the knowledge-base. *Psychonomic Bulletin and Review*, *3*, 385–388.
- Brown, N. R., & Siegler, R. S. (2001). Seeds aren't anchors. *Memory & Cognition*, *29*, 405–412.
- Brown, N. R., & Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. *Journal of Sex Research*, *36*, 292–297.
- Brunswik, E. (1955). Representative design and probabilistic theory in functional psychology. *Psychological Review*, *62*, 193–217.
- Burt, C. D. (1992). Retrieval characteristics of autobiographical memories. *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.
- Conrad, F., Brown, N. R., & Cashman, E. (1998). Strategies for answering behavioral frequency questions. *Memory*, *6*, 339–366.
- Conway, M. A. (1996). Autobiographical knowledge and autobiographical memories. In D. C. Rubin (Ed.), *Remembering our past*, (pp. 67–93). Cambridge: Cambridge University Press.
- Friedman, A., & Brown, N. R. (2000a). Reasoning about geography. *Journal of Experimental Psychology: General*, *129*, 193–219.
- Friedman, A., & Brown, N. R. (2000b). Updating geographical knowledge: Principles of coherence and inertia. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 900–914.
- Friedman, A., Brown, N. R., & McGaffey, A. (in press-a). A basis for bias in geographical judgements. *Psychonomic Bulletin and Review*.
- Friedman, A., Kerkman, D., & Brown, N. R. (in press-b). The role of proximity in spatial location judgements: A cross-national comparison of estimation bias in subjective North American geography. *Psychonomic Bulletin and Review*.
- Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, *35*, 1–36.
- Friel, S. A., & Carlson, B. W. (2000, June). Varying types of anchor primes affect numerical estimation. Poster presented at the Annual Meeting of the American Psychological Society, Miami, FL.
- Gigerenzer, G. (Ed.). (1999). *Simple heuristics that make us smart*. New York: Oxford.
- Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review*, *98*, 352–376.
- Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytical cognition: Memory, perception, and concept learning. *The psychology of learning and motivation*. (Vol. 18, pp. 1–47). New York: Academic Press.
- Jacoby, L. L., Kelley, C. M., Brown, J., & Jasechko, J. (1989). Becoming famous over night: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality and Social Psychology*, *56*, 326–338.
- Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, *21*, 1161–1166.

- Kemp, S., & Burt, C. D. B. (1998). The force of events: Cross-modality matching the recency of news events. *Memory*, 6, 297–306.
- Kerkman, D. D., Norris, K., & Stea, D. (2001). *Diversity orientation, natio-centrism, and experience: Texans' views of Mexico, Canada, and the U.S.* Manuscript in preparation.
- Kerkman, D. D., Friedman, A., Brown, N. R., Stea, D., & McCormick, A. (2001). *The development of geographic categories and biases.* Manuscript submitted for publication.
- LaVoie, N. N., Bourne, L. E., Jr., & Healy, A. F. (in press). Memory seeding: Representations underlying quantitative estimation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Lawson, R., & Bhagat, P. S. (in press). The role of price knowledge in consumer product knowledge structures. *Psychology and Marketing*.
- Lee, P. J. (2001). *The role of guessing and boundaries in the forward and backward telescoping of news events.* Unpublished manuscript, University of Alberta, Edmonton, Alberta, Canada.
- Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 551–578.
- Malmi, R. A., & Samson, D. J. (1983). Intuitive averaging of categorized numerical stimuli. *Journal of Verbal Learning and Verbal Behavior*, 22, 547–559.
- Menon, G. (1993). The effects of accessibility of information in memory on judgments of behavioral frequencies. *Journal of Consumer Research*, 20, 431–440.
- Murray, K., & Brown, N. R. (2001). *Estimating and updating internal reference prices: An examination of the structure of knowledge for automobile prices.* Manuscript in preparation.
- Nickerson, R. (1980). Motivated retrieval from archival memory. In J. H. Flowers (Ed.), *Nebraska Symposium on Motivation 1980* (pp. 73–119). Lincoln: University of Nebraska Press.
- Norman, D. A., & Bobrow, D. C. (1979). Descriptions: An intermediate stage in memory retrieval. *Cognitive Psychology*, 11, 107–123.
- Parducci, A. (1963). Range-frequency compromise in judgment. *Psychological Monographs*, 77 (2, Whole No., 565).
- Paulos, J. A. (1990). *Innumeracy: Mathematical illiteracy and its consequences.* New York: Vantage.
- Poulton, E. C. (1989). *Bias in quantifying judgments.* Hillsdale, NJ: Erlbaum.
- Reisser, B. J., Black, J. B., & Abelson, R. P. (1985). Knowledge structures in the organization and retrieval of autobiographical memories. *Cognitive Psychology*, 17, 89–137.
- Shum, M. S. (1998). The role of temporal landmarks in autobiographical memory processes. *Psychological Bulletin*, 124, 423–442.
- Spencer, J. (1961). Estimating averages. *Ergonomics*, 4, 317–328.
- Strack, F., & Mussweiler, T. (1997). Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility. *Journal of Personality and Social Psychology*, 73, 437–446.
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers.* San Francisco, Jossey-Bass.
- Thompson, C. P., Skowronski, J. J., & Betz, A. (1993). The use of partial temporal information in dating personal events. *Memory & Cognition*, 21, 352–360.
- Thompson, C. P., Skowronski, J. J., Larsen, S. F., & Betz, A. (1996). *Autobiographical memory: Remembering what happened when.* Mahwah, NJ: Erlbaum.
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 4, 207–232.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- Walbaum, S. D. (1997). *Seeding the fast food knowledge-base: Long-term effects.* Poster Presented at the 38th Annual Meeting of the Psychonomic Society, Philadelphia, PA.
- Williams, M. D., & Hollan, J. D. (1981). The process of retrieval from very-long term memory. *Cognitive Science*, 5, 87–119.