Psychological Review
1993, Vol. 100, No. 3, 511–534

0033-295X/93/$3.00

Metrics and Mappings: A Framework for Understanding Real-World Quantitative Estimation

Norman R. Brown and Robert S. Siegler

Estimation is influenced by a variety of processes: application of heuristics, domain-specific reasoning, and intuitive statistical induction, among them. In this article, we propose the metrics and mapping framework to account for how these processes are integrated to generate estimates. This framework identifies 2 types of information as critical: knowledge of distributional properties (metric knowledge) and knowledge of relative status of individual entities within the distribution (mapping knowledge). Heuristics and domain-specific knowledge are both viewed as cues that contribute to mapping knowledge; intuitive statistical induction is viewed as providing cues to metric properties. Results of 4 experiments illustrate the framework's usefulness for integrating these types of information and for predicting when people emphasize heuristics and when they emphasize domain-specific knowledge.

Quantitative estimation is a process through which people assign numerical values to properties of objects, events, and abstractions. From the estimator's perspective, these values are estimates because the true value is unknown, unknowable, or arbitrary. Psychologists have asked people to estimate many types of values. Some exist only within the context of the experiment: for example, the loudness of a tone, the brightness of a light, or the length of a line presented in the laboratory. Others involve more enduring properties of the external world: for example, the distances between cities, the dates of occurrence of public events, or the likelihood of dying by particular causes.

This latter type of estimation, real-world quantitative estimation, is the focus of the present article. It is important both because of its pervasiveness and because of its centrality for understanding the social and physical environments. We live in a world of quantitative dimensions. Reasonably accurate estimation of quantitative values is necessary for understanding that world. Without being able to estimate that the universe was millions or billions rather than a few thousand years old, Darwin could not have formulated the theory of evolution. Without being able to estimate that a 1992 U.S. government deficit of $320 billion implies a debt of about $3,000 per family, rather than $300 or $30, it is impossible to evaluate current economic policy. Without being able to estimate the time required to drive to work, we would not even know when to leave home in the morning.

Previous research has identified a number of factors that influence estimation. Some have been labeled *heuristics*; others others have been labeled *domain-specific knowledge*. A great deal of evidence has been marshaled, demonstrating that each type of knowledge influences estimation in a wide variety of contexts. However, the research has not culminated in any theory of estimation, not even in a coherent framework for thinking about the process. This gap is reflected in the strangely bifurcated nature of research in the area. Research on heuristics does not indicate when, if ever, estimation is also influenced by domain-specific knowledge; research on domain-specific knowledge does not indicate when, if ever, estimation is also influenced by heuristics. Because of this neglect, the critical issues of when people rely most heavily on heuristics, when they rely most heavily on domain-specific knowledge, and how these (and other) types of knowledge are integrated to arrive at estimates remain poorly understood.

This problem of our ignorance concerning the factors governing relative reliance on heuristics and domain-specific knowledge has not been specifically addressed previously. However, a special case of the problem has often been noted and lamented: our ignorance concerning the factors governing reliance on particular heuristics. Consider the following comments, appearing in articles over the past 15 years:

> Heuristics may be faulted as a general theory of judgment because of the difficulty of knowing which will be applied in any particular instance. (Slovic, Fischhoff, & Lichtenstein, 1977, p. 6)

> It is difficult to determine which heuristic applies to a situation and the particular manner in which it applies. (Wallsten & Barton, 1982, p. 362)

> There are few guidelines to tell us when and under what conditions each heuristic will be applied. (Sherman & Corty, 1984, p. 191)

> Our problem, then, is to describe as definitely as possible those tasks or environmental conditions that induce, or even permit, specific nonextensional heuristics to occur. . . . Otherwise, re-

---

Norman R. Brown, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada; Robert S. Siegler, Psychology Department, Carnegie Mellon University.

This research was supported by National Institute of Mental Health Training Grant (MH19102), which supported Norman R. Brown's postdoctoral training, and by research grants from the Mellon Foundation and the National Institutes of Health (HD19011).

We thank Cathy Dennler for her assistance in running experiments and Daniel Kahneman and John Kihlstrom for helpful reviews.

Correspondence concerning this article should be addressed to Robert S. Siegler, Psychology Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, or to Norman R. Brown, Department of Psychology, Biological Sciences Building, University of Alberta, Edmonton, Alberta, Canada T6G 2E9.

> search on cognitive illusions offer only a list of heuristics that have been demonstrated to occur under conditions created to demonstrate them. (Hammond, 1990, p. 243)

Gigerenzer, Hoffrage, and Kleinbolting (1991), Pitz and Sachs (1984), and numerous others have made similar observations. Despite this widespread recognition of the problem, however, solutions have not been forthcoming.

Our goal in the present article is to propose one solution, a framework for integrating the diverse processes involved in real-world quantitative estimation. We have labeled it the *metrics and mappings framework*. In the sections that follow, we first review findings generated by three approaches relevant to estimation: approaches emphasizing heuristics, reasoning from domain-specific knowledge, and intuitive statistics. We then present the metrics and mappings framework and discuss how it can be used to integrate the three approaches and to predict when and for what purposes each process will be most relevant. After this, we report four experiments designed to test, refine, and demonstrate the framework's usefulness for understanding estimation in general and estimation of national populations and areas in particular. Finally, we consider several general implications of the analysis and findings, including theoretical implications regarding how diverse processes work together and educational implications concerning how estimation can be improved.

Three Approaches to Real-World Quantitative Estimation

Heuristics

One approach to estimation is based on the assumption that people use general purpose heuristics such as availability, representativeness, and anchoring to estimate quantitative properties (Brown, Rips, & Shevell, 1985; Combs & Slovic, 1979; Fischhoff, 1987; Hogarth, 1987; Slovic, Fischhoff, & Lichtenstein, 1982; Smith & Kida, 1991; Tversky & Kahneman, 1973, 1974). For example, a person attempting to date Oliver North's dismissal from the National Security Council might realize that he or she knew a lot about the event. This would imply, through application of the availability heuristic, that North's dismissal was relatively recent. In the domain of greatest interest within this article, population estimation, a person might reason, "I know almost nothing about Indonesia; I've hardly ever heard of it; its population must be small."

Consistent with this perspective, real-world estimates have been found to be systematically biased in the direction predicted by the heuristics. For example, people tend to consider well-known events to be more recent than less well-known events of the same objective age (Brown et al., 1985; Means, Nigam, Zarrow, Loftus, & Donaldson, 1989; Wagenaar, 1986). They also estimate well-publicized causes of death, such as accidents, to be more probable than less publicized but more frequent causes, such as strokes (Combs & Slovic, 1979; Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978; Slovic et al., 1982).

These and related findings have been interpreted as indicating that people rely on availability and other broadly applicable heuristics for estimating values of real-world properties. For example, Tversky and Kahneman (1973) wrote,

> One may assess the divorce rate in a given community by recalling divorces among one's acquaintances; one may evaluate the probability that a politician will lose an election by considering various ways in which he may lose support; and one may estimate the probability that a violent person will "see" beasts of prey in a Rorschach card by assessing the strength of association between violence and beasts of prey. In all of these cases, the estimation of the frequency of a class or the probability of an event is mediated by an assessment of availability. A person is said to employ the availability heuristic whenever he estimates frequency or probability by the ease with which instances or associations could be brought to mind. (pp. 208–209)

Tversky, Kahneman, and other investigators who emphasize heuristics do not claim that availability, and availability alone, determines people's estimates. However, the just-cited description is typical of their depictions of the estimation process. Availability is the only influence cited, and people are said to reach estimates by relying on it. Certainly, no specific attention is given to determining whether, much less how, people integrate availability with relevant domain-specific knowledge.

Domain-Specific Knowledge

A second approach to real-world quantitative estimation depicts people as estimating real-world values by recalling and drawing inferences from specifically relevant aspects of their knowledge of the domain (Baddeley, Lewis, & Nimmo-Smith, 1978; Brown, 1990; Collins, 1978a, 1978b; Collins & Michalski, 1989; Ferguson & Martin, 1983; Friedman, 1987; Friedman & Wilkins, 1985; Linton, 1975; Means et al., 1989; Thompson, 1982). For example, a person estimating when Oliver North was dismissed might recall that it happened during Ronald Reagan's last term in office and that Reagan's last term lasted from January 1985 to January 1989. The person could then deduce that North was dismissed between 1985 and 1989. Similarly, people could estimate the population of Libya by reasoning, "Libya is almost all desert: deserts are usually sparsely inhabited; Libya probably has a small population."

Some of the clearest evidence for the influence of domain-specific knowledge comes from research on temporal estimation. When subjects are asked to recall when memorable personal or public events occurred, they often cite other events whose dates they know and then make such statements as "It happened just after that" (e.g., Baddeley et al., 1978; Brown, 1990; Friedman & Wilkins, 1985). Even when people do not know exactly when events occurred, knowing other facts about the events enhances the accuracy of their estimates (e.g., Brown et al., 1985; Means & Loftus, 1991; Thompson, Skowronski, & Lee, 1987). Some researchers who have focused on domain-specific knowledge have recognized that "nonverbal" processes, which in the context appeared to refer to heuristics, may influence estimation (e.g., Collins & Michalski, 1989). However, they have not specified in any detail what these nonverbal processes are, nor how they affect performance. For example, although Collins and Michalski noted that nonverbal processes could influence estimates, they based their analysis of estimation exclusively on verbal comments that emphasized domain-specific knowledge such as the presence and effects of deserts, jungles, and mountains. Thus, research that emphasizes domain-specific knowledge, like research that emphasizes heuristics, does not indicate whether, much less how, the two are integrated to produce estimates.

Intuitive Statistics

A third line of research relevant to real-world estimation is based on the metaphor of man as an intuitive statistician (Chapman & Chapman, 1969; Edwards, 1968; Gigerenzer et al., 1991; Gigerenzer & Murray, 1987; Kunda & Nisbett, 1986; Peterson & Beach, 1967). Most research embodying this perspective has been concerned with inferential statistics; some research, however, has focused on how people induce descriptive statistical properties. Although this research was not conducted with real-world quantitative estimation in mind, it is nonetheless relevant to it. Part of what people need to learn to accurately estimate distances, mountain heights, populations, gross national products, and other aspects of the world are the types of numerical values that go with the domain.

A main focus in these studies of descriptive statistics has been on induction of means and other measures of central tendency (N. H. Anderson, 1964; Beach & Swenson, 1966; Hendrick & Costantini, 1970; Krueger, Rothbart, & Sriram, 1989; Levin, 1974, 1975; Malmi & Samson, 1983; Spencer, 1961, 1963). For example, Spencer (1961) reported that estimates of the mean of sets of 10 or 20 numbers were very accurate, almost always within 10% of the correct value. Results were similar when the task was to identify the median (Beach & Swenson, 1966).

Abstraction of properties other than means and medians has received less attention. However, Beach and Scopp (1968) and Lathrop (1967) have demonstrated that subjects are sensitive to the variability of different sets of numbers, and Levin (1975) has demonstrated that direct ratings of variance can be quite accurate.

The situations in which these results have been obtained differ from those involved in real-world quantitative estimation in that subjects in the intuitive statistics experiments were only presented numbers; they did not need to learn about the connection of numbers to particular entities. However, the findings suggest that people may be able to abstract numerical properties from encountering numbers attached to objects and events as well as from encountering the numbers in isolation.

Combined Influences

Although these three approaches have developed separately, all are relevant to understanding estimation. The processes they emphasize seem far more likely to jointly influence estimates than to operate in isolation. Preliminary evidence for this assertion can be found in existing data sets. For example, Brown et al. (1985, Experiment 1) reported a correlation of r = .88 between true and estimated dates of occurrence of a set of public events. When availability (as measured by subjects' ratings of their knowledge of the events) was partialed out, the correlation remained significant, r = .90; apparently, specific knowledge of the events allowed subjects to produce quite accurate estimates above and beyond any influence of the availability heuristic. Conversely, when the objective dates were partialed out, the correlation between estimated date and availability also remained significant, r = .54; apparently, the availability heuristic also influenced the estimates. The fact that these findings arose from a single group of subjects performing a single task within a single experiment leaves little doubt that both availability and domain-specific knowledge influenced the estimates.

We believe that any satisfactory model of quantitative estimation must account for the influence of heuristics, domain-specific qualitative knowledge, and statistical knowledge as well. In the next section, we describe a conceptual framework for integrating the processes described by the three approaches and for understanding their unique and combined contribution.

Metrics and Mappings Framework

General Perspective

Each of the three approaches described earlier focuses on a single type of process that is frequently involved in estimation: (a) application of heuristics, (b) conscious explicit reasoning, and (c) induction of numerical properties. Within each approach, the individual process is the focus of interest.

Our research has a somewhat different purpose: to understand estimation as a whole. We believe that estimation involves all three types of processes (and probably others as well). For this reason, we believe it can be understood only by establishing the roles of each component process and determining how they jointly determine people's estimates. This perspective has led us to a research strategy like that suggested by Newell's (1973) recommendation that researchers attempt to account for "a genuine slab of human behavior" (p. 303) rather than focusing on a given process in isolation.

The two research strategies, focusing on a specific process and focusing on the set of processes that generate the behavior of interest, are complementary rather than antagonistic. The more that is known about each component process, the more precise an integrative model can be. The more precise the integrative model, the more penetrating the questions that can be asked about the components. Nonetheless, the two research strategies also lead to different questions and different empirical research. For example, neither approaches that emphasize heuristics nor approaches that emphasize domain-specific reasoning have focused on the absolute numerical characteristics of estimates. In the one case, the emphasis is on the direction and degree to which estimates deviate from correct values; in the other case, the emphasis is on how people reason from consciously available knowledge. However, for an approach aimed at explicating how a subject arrives at a particular numerical estimate, absolute as well as relative magnitudes of estimates are basic data to be explained. The metrics and mappings approach is aimed at accounting for both.

Central Conceptual Distinction

Figure 1 represents a basic framework for thinking about real-world quantitative estimation. The key distinction within it analyzes estimation into metric and mapping components.

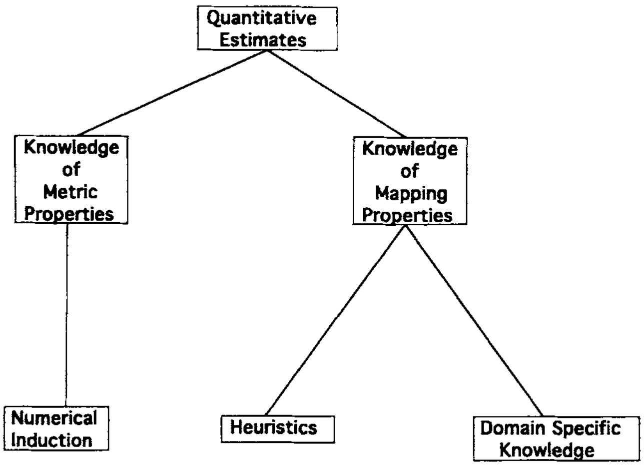

Figure 1. Basic structure of metrics and mappings framework.

Metric knowledge is believed to derive primarily from numerical induction, a process focused on in the intuitive statistics approach; knowledge of mappings is believed to derive primarily from domain-specific reasoning and application of heuristics. Thus, the framework indicates how the three literatures described earlier contribute to understanding of different facets of estimation, as well as indicating the types of processes believed to generate each type of knowledge.

The central idea within this framework is that real-world quantitative estimation requires two types of information: metric knowledge and mapping knowledge. Metric knowledge focuses on statistical properties of the domain, such as the mean, median, variance, and form of the distribution. For example, to accurately estimate the career batting averages of 10 randomly chosen major league baseball players, it helps to know that averages tend to be around .250, that they rarely exceed .330 or fall below .200, and that most players other than pitchers bat between .220 and .280. Such metric information tells us about plausible numerical values of the measure and constrains us from making estimates that are totally "out of the ballpark."

In contrast, mapping components involve ordinal relations among individual entities within the domain. Often, people who have no metric knowledge possess ordinal information that allows them to map newly presented metric information onto appropriate individuals. For example, a casual baseball fan who had often heard of Barry Bonds, occasionally of Jay Bell, and rarely or never of Gary Varsho might have little idea what their batting averages were. When told that last year one player hit .300, one .260, and one .220, however, the fan would have a good guess concerning which player hit for which average.

At first, the distinction between metric and mapping properties might seem the same as that between cardinal and ordinal properties (or between absolute and relative properties). The distinctions differ considerably, however. Metric properties, such as the mean, median, and variance, describe the distribution, not the individual entities within it. Knowing such properties implies nothing about the relative status of individuals. Mappings, in contrast, specify the ranking of individuals on the dimension of interest but say nothing about the distribution per se. Seen from another perspective, complete metric knowledge can be represented by a scatterplot containing unnamed points that specify all values within a distribution. Complete mapping knowledge can be represented by a list containing the rank of each individual entity along the relevant dimension. As can be seen by imagining mapping the ordered list onto the scatterplot, metric and mapping properties together completely specify the information that needs to be estimated, but neither alone does.

The distinction between cardinal and ordinal values differs from the metrics and mappings distinction in at least two critical ways. First, both cardinal and ordinal values characterize individual entities within a distribution, rather than the distribution as a whole, which is the basic unit for metric knowledge. Second, cardinal and ordinal values are not independent; the cardinal values of two or more entities imply their ordinal relation, whereas there is no implicative relation between metric and mapping properties.

Figure 2 illustrates the conceptual independence of metric and mapping information. Panel A represents a base case: Nielsen Poll data concerning the average amount of television watched per week by people of different ages (Information Please Almanac, 1990). Panels B, C, and D are hypothetical estimates of these data. Panel B represents estimates with different metric properties than those in Panel A but the same mappings (i.e., the same ordinal relations between age and viewing time); Panel C represents estimates with different mappings than those in Panel A but identical metric properties (i.e., the same distribution of numbers, but with different numbers attached to each age range); Panel D represents estimates different in both ways.

The conceptual distinction is also evident in the ways in which new information is likely to influence estimates. Suppose a person needed to estimate the weekly mean number of hours of TV watched by people in the six age groups in Figure 2. Saying to the person who generated the estimates in Panel B "Your estimates are too low" would probably lead the person to change the numerical distribution of estimates (by raising them) but leave unchanged the ranking of age groups within the distribution. Such information would change the metric but not the mapping properties of the estimates. Conversely, telling the person who generated the estimates in Panel C "Children and teenagers actually watch less TV than adults" would probably result in increased estimates for adults and decreased ones for children and teenagers. However, there would be little reason to expect systematic changes in the mean or median of the estimates. Thus, the second statement would affect mapping properties but not metric ones. Data reported in this article demonstrate that new information often does exert entirely different effects on metric and mapping properties.

Measurement Implications

The cost of not differentiating between metric and mapping components is reflected in the deficiencies of one of the most commonly used measures of estimation accuracy: mean absolute error. This measure is calculated by

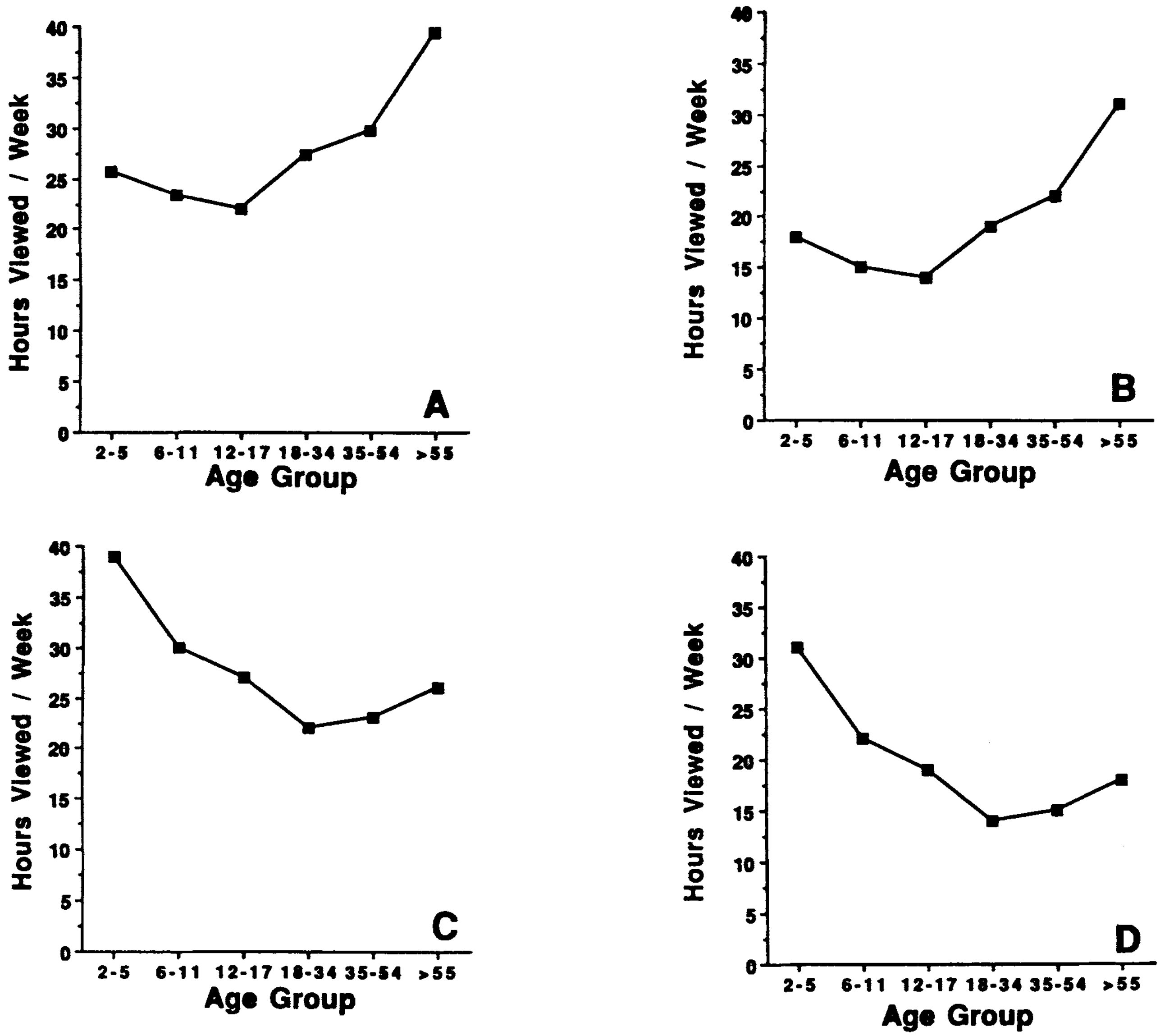

Figure 2. Mean hours of television watched per week by different age groups. (Panel A: actual data from Information Please Almanac, 1990; Panel B: transformed data with same mappings but different metric properties; Panel C: transformed data with different mappings but same metric properties; Panel D: transformed data with different mappings and different metric properties.)

$$\frac{\Sigma_{ij}|E_{ij} - T_j|}{IJ},$$

where E_{ij} is a particular subject's estimate on a particular item, T_j is the true value for that item, and IJ is the number of observations. For example, if a subject estimated the mean television viewing time of two age groups, and one estimate was 5 hr too high and the other 5 hr too low, the mean absolute error would be 5.

The problem with this measure is that it reflects an unknown mixture of two types of errors: error that is due to incorrect metric properties and error that is due to incorrect mappings. Figure 2 is again illustrative. Relative to the true values in Panel A, the estimates in Panels B and C yield identical mean absolute errors of 8 hr. However, the source of the error is entirely different. The estimates in Panel B are uniformly too low, whereas those in Panel C are correct in all metric properties but incorrect in their mappings of age groups onto viewing time. If we wanted to improve the estimates in Panels B and C, we would presumably provide the estimator with entirely different information. Absolute error can be useful for conveying a general sense of the degree of inaccuracy of estimates, but it is ambiguous as to the source of the inaccuracy.

In contrast, the distinction between metric and mapping components directly suggests a pair of more differentiated measures of the accuracy of estimates: median overall deviation (MOD) and rank-order correlations. MOD is a pure measure of a metric property. It is calculated

$$\frac{\Sigma_j |Med_j - Med_t|}{J},$$

where Med_j is the median of each subject's estimates, Med_t is the true median, and J is the number of subjects. As suggested by this formula, MOD depends only on the discrepancy between the median estimate of each subject across all items and the true overall median, not at all on the correlations between the estimated and true value for each item. The lower the value of MOD, the more accurate the central tendency of that subject's estimates. Returning to Figure 2, MOD for the estimates in Panel B is 8, because each numerical value (and therefore the median of the values) deviates from the true value shown in Panel A by this amount. In contrast, MOD for the Panel C estimates is 0, because the six numerical values (and therefore their median) are the same as the true values in Panel A.[1]

Conversely, rank-order correlations provide a measure of mapping independent of metric properties of the estimates. Returning to Figure 2, the rank-order correlation between the Panel A and Panel B data is 1.00. However, the rank-order correlation between the Panel A and Panel C data is −.60.

Together, the two measures, MOD and the rank-order correlation, indicate the source of the difficulty of the estimates represented in each panel. MOD indicates that the Panel B estimates are too high; the rank-order correlation indicates that the Panel C mapping between age group and amount of viewing is incorrect. Thus, the two measures differentiate between the two sources of inaccuracy.

Implications Concerning What Data Are Relevant

Previous studies of real-world quantitative estimation have minimized attention to knowledge of absolute quantities. Sometimes they have asked subjects to use arbitrary response ranges in which metric properties have little meaning (e.g., "Rate the extensiveness of the auditing controls you would do to detect this type of error"; Smith & Kida, 1991). Other times, they have asked only for relative judgments (e.g., "Are homicides more likely than suicides?"; Lichtenstein et al., 1978). Still other times, the experimenter provides important metric information to the subjects (e.g., "All events happened between 1976 and 1983"; Brown, 1990). Even when absolute estimates have been available and meaningful, they generally have been reported only in passing or not at all (e.g., Tversky & Kahneman, 1973).

The present framework, in contrast, suggests that metric properties as well as mappings are important. Their importance is not just theoretical but practical as well; real-world events frequently cannot be understood without them. Even ratios of two estimated values, which have been reported in some previous experiments, do not convey the same information as absolute quantities. It is one thing to know that the U.S. deficit each year is 10 times as big as it was a relatively few years ago; it is quite another to know that in 1992, it was roughly $3,000 rather than $300 per family (as opposed to $30 vs. $3). Many real-world decisions simply cannot be made rationally without reasonable estimates of the quantities involved.

Integrating Alternative Approaches to Estimation

Once we divide estimation into metric and mapping components, we can examine the processes most prominent in generating each type of information. Metric information seems likely to come primarily from the type of induction of numerical properties that has been studied within the intuitive statistics approach. There simply is no substitute for the actual numbers if the goal is to induce central tendencies, variability, and other metric characteristics.

Mappings, however, can be derived both from application of heuristics and from reasoning based on domain-specific qualitative knowledge.[2] In the context of the present study, both availability (a broadly applicable heuristic) and size on a map (a factor specific to the domain of geographic entities) suggest orderings of different countries' populations. Neither the heuristic nor the domain-specific information usually indicates the relevant metric properties (e.g., whether populations are generally in the hundreds of thousands, millions, or billions), but both often are useful for establishing ordinal relations among the populations of different countries.

This framework makes clear that heuristics and domain-specific knowledge are used to reach a common goal. The framework also suggests a way to reconcile approaches that emphasize the two types of information. Traditionally, heuristics and domain-specific knowledge have been thought of as distinct types of entities. However, it may be more profitable to think of both in the same way: as cues with varying validities in different domains that can be used to predict mappings of quantitative values onto real-world entities. From this perspective, heuristics may simply be names we give to cues that are predictively useful in many situations.

Suppose that we view both heuristics and domain-specific knowledge as cues rather than as qualitatively different types of entities. What implications does this have for our perspective on estimation? One implication is that the same type of cue-validity analyses that have proven useful in studying many other aspects of perception, cognition, and language (e.g., J. R. Anderson, 1990; Brunswik, 1956; Gigerenzer et al., 1991; Huttenlocher, Hedges, & Duncan, 1991; MacWhinney, 1987; McClelland & Rumelhart, 1986; Mervis & Rosch, 1981) will also prove useful in understanding real-world quantitative estimation. In contrast with prior approaches to estimation, approaches emphasizing cue validities suggest that (a) estimates are a weighted blend of different sources of information and (b) the relative weighting of each source reflects its predictive strength relative to the predictive strengths of competing sources.

---

[1] The reason that medians rather than means were used to calculate MOD was practical rather than theoretical. In real-world quantitative estimation, estimates often span several orders of magnitudes. When means are used, outlying estimates can have enormous effects; these effects are reduced when medians are used.

[2] Domain-specific quantitative information can yield knowledge about both metric and mapping properties. For example, learning that the per capita incomes of Brazil, Chile, Ghana, and Kenya are $1,976, $1,330, $380, and $230, respectively, communicates information both about metric properties of third-world incomes and about per capita incomes in South America generally being higher than those in Africa.

As has become clear within research emphasizing cue validities in other areas, the predictive accuracy of a cue is not the only factor determining its weighting. For example, MacWhinney, Leinbach, Taraban, and McDonald (1989) found that reliance on a cue is influenced not only by its predictive validity but also by its availability, encodability, and specific validity when it is in conflict with other cues (its conflict validity). Similarly, Kunda and Nisbett (1986) found that both familiarity and encodability of the data influenced subjects' success in identifying correlations. Consideration of all of these factors, rather than of predictive validity alone, is likely to yield the best estimates of relative weightings. Among cues that are familiar and likely to be encoded, however, relative predictive validity seems to provide a good rule of thumb for anticipating relative weightings.

This assumption leads to a prediction concerning when heuristics should exert the greatest influence on estimates and when domain-specific knowledge should. On the one hand, in situations where most people's domain-specific knowledge is limited or not highly predictive, heuristics such as availability should be most influential. On the other hand, in domains where people do have relevant, domain-specific knowledge that has higher cue validity than available heuristics, the domain-specific information should be emphasized. This perspective implies that the type of cue that receives the greatest weighting is not a function of heuristics generally being more influential or of domain-specific knowledge generally being more influential: The relative weighting is instead a function of the relative predictive value of available information in the particular domain.

Estimating National Populations

The four experiments described in this article all required subjects to estimate national populations. This task was of interest for a number of reasons. One was that it represents an extremely common but rarely studied class of situations in which people estimate: situations where we possess little or no quantitative data but considerable relevant qualitative information. A pilot study indicated that most American college students know the populations (within 10%) of few or no countries. Yet they often do possess both general heuristics and knowledge of specific facts that provide a basis for estimating the populations. For most people, then, population estimation requires use of a variety of qualitative cues to draw a quantitative conclusion.

Beyond providing an opportunity to study estimation in situations where qualitative information is rich and quantitative information poor, population estimation allowed us to address several issues of specific theoretical interest. It enabled us to test the utility of the central distinction between metric and mapping components, both by examining whether specific manipulations affected the two components differently and by testing whether different facets within each component responded similarly to a given manipulation. Also, because the number of nations is finite and relatively small, we could compute predictive validities for several cues and draw contrasting predictions concerning the relative weighting of heuristics and domain-specific knowledge for population and area estimation. The domain also provided an opportunity to test a research method that may prove generally useful for studying estimation. This method involves *seeding the knowledge base* with a small number of quantitative values from a domain and observing the impact on subsequent estimates of encountering other values in it. As shown in Experiments 2–4, results obtained from this approach illustrate the importance of the basic conceptual distinction between metric and mapping components. Some sets of seed facts affect just the metric component, others affect just the mapping component, others affect both, and others affect neither. Which seed facts affected which component (or components) could be predicted straightforwardly from the theoretical framework.

Experiment 1

In Experiment 1, subjects estimated the populations and areas of 99 of the 100 most populated countries (all except the United States, whose population and area were stated as an example in the instructions). The experiment was designed with two goals in mind. One was descriptive. Because there was no previous research on population estimation, and almost none on land area estimation, it seemed essential as a first step to obtain empirical data about people's existing knowledge. This would inform us about both metric and mapping properties of estimates in these domains.

The second goal was to examine how cue validities influence estimation. Few people have precise quantitative information about national populations and land areas. In the absence of such information, we expected subjects to make use of both heuristics, such as availability, and of domain-specific knowledge, such as images of maps and globes.

We made two main predictions. The first was that a domain-specific predictor, true land area, would better predict estimates of land area than would either availability or true population. The second was that a heuristic, availability, would better predict estimates of population than would either true population or true land area. The reasons for these predictions involve both cue validity and cue prevalence.

First consider predictors of estimated land area. Relative size on maps provides a familiar and highly valid cue to relative land areas. People are quite skillful in using this cue. Kerst and Howard (1978) presented subjects with a map of the United States, then withdrew it, and then asked them to estimate the ratios of different states' areas. They reported a correlation of $r = .89$ between estimated and actual ratios of the areas. The current situation differed from that studied by Kerst and Howard in that no maps were presented in our experimental situation. However, we expected that prior experience with maps would allow subjects to have learned the general validity of the cue and to have obtained fairly accurate information about nations' relative areas.[3] In contrast, availability is not nearly as

[3] This assumption depends on subjects knowing which countries go in particular positions on the map. Our sample appears to have met this assumption quite well. When we asked a comparable Carnegie Mellon University undergraduate sample to locate 16 relatively well-known countries on the map (e.g., France and Pakistan), they averaged more than 15 correct.

predictive of land area; an a priori estimate of each country's availability, its number of citations in the 1988 New York Times Index, correlated only $r = .19$ with the country's land area for the 99 countries studied in Experiment 1. This led us to predict that estimates of land area would be better predicted by true land area than by availability.

The situation was reversed with population estimates. Whereas globes and world maps are a familiar feature of classrooms and homes, information about populations is much less prevalent in the everyday environment. Populations can be estimated on the basis of land area (larger countries tend to have more people), but our assessment of ecological validities indicated that availability is a better predictor. Over the 99 countries in the assessment, the rank-order correlation between true population and citations in the 1988 New York Times Index was .56; the rank-order correlation between true population and true land area was .47. These considerations of cue validity and cue prevalence indicated that availability should be a better predictor of estimated population than should true land area, whereas true land area should be a better predictor of estimated land area than should availability.

Method

Materials. In 1989, exactly 100 countries were believed to have populations of 4 million or more (Information Please Almanac, 1989). These countries, except for the United States, served as the test items in Experiment 1. The 99 countries (listed in Table 1) had a mean population of 48.3 million, a median population of 15.1 million, a mean area of 450,000 square miles, and a median area of 127,000 square miles. As these statistics suggest, the distributions of national populations and land areas are highly skewed, with many small countries and a few large ones.

Design and procedure. At the beginning of Experiment 1, subjects were given separate booklets for three tasks: population estimation, area estimation, and knowledge rating. Task order was counterbalanced, with 25% of subjects performing the tasks in each of four orders: PAK, PKA, KAP, and KPA (where P stands for population estimates, A for area estimates, and K for knowledge ratings). Each booklet contained names of the 99 test countries, listed in a unique random order for each subject and for each task, with 20 countries listed per page. A prompt ("population," "land area," or "knowledge") and a blank line were printed under the name of each country.

In the population estimation instructions, subjects were asked to estimate the current population of each of the 99 test countries. The instructions noted that all countries were among the 100 most populated in the world and that the current population of the United States was 246.1 million. Subjects were instructed to work through the booklet at their own pace, not skipping any countries, and to make their best guess when they were uncertain.

Area estimation instructions differed in only two ways. First, subjects were informed that the land area of the United States was 3,615,000 square miles. Second, there was nothing analogous to the statement that the test items were among the 100 most populous countries in the world (because they were not all among the 100 countries having the largest areas).

Knowledge ratings provided a measure of each country's availability for each subject. Subjects were told to rate their knowledge of each test country on a 0 to 9 scale, writing 0 when they knew nothing about a country, 9 when they knew a great deal about it relative to what they knew about other countries, and 1 to 8 to represent intermediate degrees of knowledge.[4]

Subjects. Twenty-four Carnegie Mellon University undergraduates participated, all for course credit. The relatively high Standard Achievement Test (SAT) scores at this university (mean SAT Math + Verbal = 1225) suggested that their mean ability to estimate these data would be at least average for college students. Subjects were run individually or in small groups, in sessions that lasted about 45 min. The experiment was conducted in the fall of 1989.[5]

Results and Discussion

Overview. The median population estimate for each country is provided in Table 1. (A variety of other data, including median estimated area, New York Times Citation Index counts, and mean knowledge ratings for each country are provided in the Appendix.) As shown in Table 1 and the Appendix, many of the estimates were quite inaccurate. The fact that these were medians, where extreme values have little impact, suggests that the numbers, if anything, understate the degree of inaccuracy. Figure 3, depicting the entire distribution of estimates for two countries, Indonesia and Canada, may better communicate the extent of the inaccuracy.

To provide a summary measure of the accuracy of estimates, we computed means of each subject's mean absolute error. As noted in the introductory paragraphs, the fact that this measure is influenced by both metric and mapping properties limits its usefulness for indicating the source of inaccuracies. However, the fact that it reflects both influences also makes it useful for conveying an overall sense of the accuracy of estimates.

Analyses revealed that mean absolute error was quite large for estimates of both population and area. For population estimates, it was 19.2 million (ranging from 9.3 million to 70.8 million for different subjects). For area estimates, it was 220,000 square miles (ranging from 56,000 to 458,000 square miles). These errors often exceeded the quantity being estimated.

Another way to measure error for estimates is

$$|\log_{10}(\text{estimated value/true value})|.$$

This measure describes the discrepancy between true and estimated values, with 0 indicating that estimates were perfect and 1 indicating that estimates were off by an order of magnitude (Nickerson, 1981). Averaging over countries and then subjects, the mean values were .74 for populations and .70 for land area. For a country with 20 million people, a deviation of 74% of an order of magnitude on the high side would lead to an estimate of 109.9 million; a comparable deviation on the low side would lead to an estimate of 3.6 million.

Accurate estimates of individual populations and areas were rare. For both population and area tasks, fewer than 6% of

[4] Results of Brown and Siegler (1992) have provided evidence for the validity of this measure of availability. Population estimates were collected from two groups of subjects tested 17 months apart (before and after the Gulf War). When availability (as measured by rated knowledge) increased over time, population estimates also increased; when availability decreased over time, population estimates also decreased.

[5] The date when each experiment was run is included because populations, areas, and absolute levels of social scientific knowledge *all* vary over time.

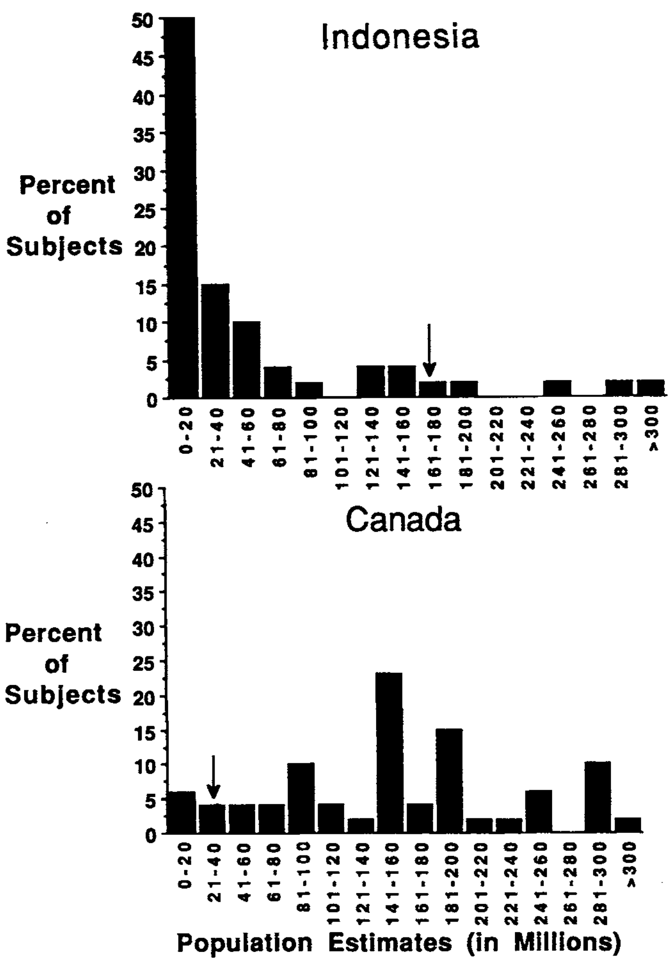

Figure 3. Distribution of estimates for Indonesia and Canada in Experiment 1. (Arrows indicate actual population for each country.)

estimates fell within 10% of the correct value. These results, in conjunction with the absolute error data, indicate that retrieval of population and area data is rarely an option, even for undergraduates at a selective U.S. university.

Metric properties. MOD, the mean of the difference between each subject's median estimate for the 99 countries and the true median for those countries, provided a pure measure of a metric property of the estimates. It proved to be quite large for both population and area estimates (mean MOD = 16.4 million people and 220,000 square miles). These errors were larger than the true medians of 15.1 million people and 127,000 square miles (Table 2).

Mapping properties. The quality of mappings between countries and their populations or areas, independent of metric properties of the estimates, could be measured by the rank-order correlation between true and estimated values. These analyses revealed moderate correlations for both population and area. The median of the individual subjects' rank-order correlations between estimated and true population for each country was .42 (range = .19 to .63). The corresponding median correlation for land area was .47 (range = .13 to .74).

Predictors of estimates. Our analysis of cue validities suggested that the best predictors of estimates would differ on the two tasks. Availability, as measured by rated knowledge of each country, was expected to be the best predictor of population estimates. In contrast, true area was expected to be the best predictor of estimated area. Thus, in one case, a heuristic would be the best predictor; in the other case, a type of domain-specific knowledge would be.

To test these predictions, we constructed separate regression models for median estimated population and median estimated land area. Each model involved regressing estimates for 96 countries against true population, true land area, and mean rated knowledge.[6]

As shown in Table 3, the predicted pattern was obtained. The best predictor of estimated area was true area; it uniquely accounted for 51% of the variance on that measure. The best predictor of estimated population was availability; it uniquely accounted for 33% of the variance on that measure. Also, the three predictors together accounted for a high percentage of the variance in both types of estimates: 81% for the area estimates and 73% for the population estimates.

Results of the regression analyses also supported in another way the assumption that heuristic and domain-specific information jointly determine estimates. In the analysis of population estimates, where availability accounted for the greatest percentage of unique variance, true area accounted for a substantial additional amount (11%). Conversely, in the analysis of area estimates, where the domain-specific predictor of true area contributed the greatest amount of unique variance, the availability heuristic accounted for a far from trivial additional amount (12%). These results again indicated that estimates of both population and area were influenced by both heuristics and domain-specific knowledge.

These results supported two assumptions of the present framework. One was that estimates are typically a weighted blend of available information, including both heuristics and domain-specific knowledge. Consistent with this assumption, measures of both availability and true land area accounted for significant independent variance in estimates of both populations and land areas.

The results also were consistent with a second assumption: Relative weighting of a cue depends on its ecological validity relative to that of alternative cues. This assumption, combined with the assessment of ecological validities, led to the prediction that for purposes of estimating population, subjects would give greater weight to availability; for purposes of estimating area, they would give greater weight to information about true

[6] The three most populous countries—China, India, and the Soviet Union—were excluded from the regression analyses. The reason was that their populations were so large that including them distorted the entire pattern. To illustrate, when the three countries are excluded, true population accounts for only 19% of the variance in estimated population; including them raises the percentage to 79%. The latter, seemingly impressive figure masks subjects' modest knowledge of the populations of the other 96 countries. For this reason, all regression analyses of population and area estimates throughout this article exclude the three extremely populous countries.

Table 1
Median Population Estimate (in Millions) for Each Country (Experiments 1 and 2)

Rank	Country	Actual population	Median estimated population in Experiment 1	Median estimated population in Experiment 2
1	China	1,087.0	855	800
2	India	816.8	225	200
3	U.S.S.R.	286.0	290	300
4	Indonesia	177.4	23	24
5	Brazil	144.4	53	60
6	Japan	122.7	100	100
7	Nigeria	111.9	17	17
8	Bangladesh	109.5	16	9
9	Pakistan	107.5	30	35
10	Mexico	83.5	100	100
11	Vietnam	65.2	19	30
12	Philippines	63.2	27	20
13	West Germany	61.2	50	60
14	Italy	57.3	42	60
15	United Kingdom	57.1	65	120
16	France	55.9	70	60
17	Thailand	54.7	13	20
18	Egypt	53.3	22	36
19	Turkey	52.9	34	22
20	Iran	51.9	18	30
21	Ethiopia	48.3	18	30
22	South Korea	42.6	34	30
23	Burma	41.1	9	10
24	Spain	39.0	50	50
25	Poland	38.0	50	50
26	South Africa	35.1	55	60
27	Zaire	33.3	11	23
28	Argentina	32.0	30	40
29	Colombia	30.6	27	25
30	Canada	26.1	150	150
31	Morocco	25.0	10	20
32	Tanzania	24.3	4	10
33	Algeria	24.2	9	20
34	Sudan	24.0	16	15
35	Yugoslavia	23.6	25	25
36	Kenya	23.3	18	15
37	Rumania	23.0	19	15
38	North Korea	21.9	27	25
39	Peru	21.3	18	16
40	Taiwan	19.8	22	20
41	Venezuela	18.8	17	20
42	Nepal	18.3	7	10
43	Iraq	17.6	23	30
44	Malaysia	17.0	12	20
45	East Germany	16.6	30	35
46	Sri Lanka	16.6	5	9
47	Australia	16.5	62	60
48	Uganda	16.4	10	20
49	Czechoslovakia	15.6	27	30
50	Mozambique	15.1	4	10
51	Netherlands	14.7	17	24
52	Afghanistan	14.5	14	35
53	Ghana	14.4	10	15
54	Saudia Arabia	14.2	33	34
55	Chile	12.6	19	25
56	Syria	11.3	20	22
57	Ivory Coast	11.2	9	8
58	Madagascar	10.9	6	8
59	Hungary	10.6	30	28
60	Cameroon	10.5	7	4
61	Cuba	10.4	6	8
62	Portugal	10.3	15	19
63	Ecuador	10.2	12	11
64	Greece	10.1	25	35
65	Belgium	9.9	11	29

Table 1 (*continued*)

Rank	Country	Actual population	Median estimated population in Experiment 1	Median estimated population in Experiment 2
66	Zimbabwe	9.7	9	13
67	Bulgaria	9.0	10	18
68	Guatemala	8.7	10	20
69	Mali	8.7	5	3
70	Burkina Faso	8.5	4	1
71	Sweden	8.4	21	22
72	Angola	8.2	10	22
73	Somalia	8.0	6	10
74	Tunisia	7.7	7	6
75	Malawi	7.7	5	2
76	Austria	7.6	24	40
77	Zambia	7.5	17	15
78	Niger	7.2	11	10
79	Rwanda	7.1	3	2
80	Senegal	7.0	5	12
81	Guinea	6.9	5	10
82	Bolivia	6.9	11	28
83	Dominican Republic	6.9	3	4
84	Cambodia	6.7	22	15
85	North Yemen	6.7	4	5
86	Switzerland	6.6	15	39
87	Haiti	6.3	2	10
88	El Salvador	5.4	14	12
89	Burundi	5.2	8	4
90	Denmark	5.1	10	28
91	Finland	4.9	20	23
92	Chad	4.8	13	10
93	Honduras	4.8	5	11
94	Benin	4.5	3	3
95	Israel	4.4	14	23
96	Paraguay	4.4	10	13
97	Norway	4.2	23	25
98	Libya	4.0	13	30
99	Sierra Leone	4.0	2	8

area learned from maps and globes. Both predictions were borne out by the data.

Experiment 2

The results of Experiment 1 supported the popular stereotype that most U.S. students have limited social scientific knowledge. None of the measures revealed impressive knowledge of populations or areas.

Experiment 2 was designed to test whether seeding the knowledge base with specific quantitative facts would lead to more accurate generalizations concerning other countries' populations. The procedure involved subjects first estimating the populations of the same 99 countries as in Experiment 1, then learn-

Table 2
Population and Area Estimates Summed Over Countries

Experiment and measure	True median	Median estimate	Mean absolute error	MOD	Median rank-order correlation between true and estimated values
Experiment 1					
Area estimates	127[a]	261[a]	220[a]	205[a]	.47
Population estimates	15.1[b]	18.0[b]	19.2[b]	16.4[b]	.42
Experiment 2[c]					
Population estimates	15.1[b]	19.0[b]	13.0[b]	19.1[b]	.41

Note. MOD = median overall deviation.
[a] Hundreds of thousands of square miles. [b] Millions of people. [c] Data from pretest estimates.

Table 3
Predictors of Population and Area Estimates

Experiment and task	Total R^2	Unique variance accounted for: Availability	True area	True population
Experiment 1				
Area estimation	.81	.12*	.51*	.02
Population estimation	.73	.33*	.11*	.02
Experiment 2				
Population estimation	.72	.39*	.10*	0

* $p < .001$.

ing the populations of 24 of these countries, then re-estimating the populations of the full set of 99 nations (the 24 *seed countries* and the 75 *transfer countries*). The main focus was on changes from the first to the second estimate of the transfer countries' populations. There is no question that people can learn numerical facts; the issue is whether they can acquire generalizable knowledge from learning such facts.

We expected that exposure to the seed facts would lead subjects to revise their prior beliefs in appropriate directions. In particular, we expected that encountering the seed facts would (a) help subjects learn metric properties of the overall distribution, such as the mean, median, and variance; and (b) improve mappings of populations onto countries, as measured by increasing rank-order correlations between estimated and true population and by decreasing correlations between estimated population and availability.

The prediction that subjects would adjust metric properties of their estimates in the appropriate direction implied that the means, medians, and variances of individual subjects' distributions of estimates for the 75 transfer countries should move in the direction suggested by their errors for the 24 seed set countries. Thus, if the mean and median of a subject's estimates for the seed countries were too high, the subject should adjust downward estimates for the transfer countries (relative to her or his initial estimates for those countries).

The seeding procedure was also expected to lead to improved mappings, as indicated by higher correlations between estimated and true populations for the transfer countries. This would come about if subjects were able to use the seed set to correctly induce the factors that predict population size or if they were able to select appropriate seed countries to serve as quantitative reference points. For example, if subjects initially thought that Sweden was more populous than Vietnam, but learned that the opposite was true, they might then also change their initial (incorrect) view that Norway had more people than Thailand.

A related prediction was that if mappings improve, then the correlation between population estimates and our measure of availability, rated knowledge of the country, should decrease. In Experiment 1, this measure of availability correlated much more highly with estimated population than with true population (rank-order correlations of .80 vs. .44). If the seeding procedure worked as expected, it should increase the number of cues in competition with availability, by providing data for inducing new demographic generalizations. These influences would be expected to move the correlation between availability and estimated population downward toward that between availability and true population.

Method

Design. The design involved four sequential phases: knowledge rating, initial estimation, learning phase, and final estimation. The countries used in three of the four phases were the same 99 as in Experiment 1. The sole exception was the learning phase, where a subset of 24 countries was chosen as the seed set. To assure that subjects encountered populations of diverse countries, we selected as members of this seed set 6 countries for each cell of a 2 (prior knowledge: high or low) × 2 (prior estimation accuracy: high or low) factorial design. This was done to provide plausible anchors for a wide range of countries and to avoid overemphasizing either implicit positive or implicit negative feedback regarding previous performance. The seed countries are listed in Table 4.

Procedure. The knowledge rating, initial estimation, and final estimation tasks were identical to those in Experiment 1, except that pre-

Table 4
Experiment 2 Seed Countries

Category	Country
High knowledge and high accuracy	South Africa
	Spain
	Egypt
	Great Britain
	Italy
	West Germany
Low knowledge and high accuracy	Netherlands
	Venezuela
	Kenya
	Romania
	Sudan
	Argentina
High knowledge and low accuracy	Israel
	Switzerland
	Greece
	Australia
	Canada
	Vietnam
Low knowledge and low accuracy	Bolivia
	Zimbabwe
	Ivory Coast
	Chile
	Zaire
	Thailand

sentation was on a computer screen rather than in a booklet. Country names were presented individually at the center of the screen. Subjects rated their knowledge of the country or estimated its population by typing the desired number two lines beneath the country's name. When satisfied that the displayed response was the one they wanted, subjects pressed the "Enter" key, which cleared the screen. After a 0.5-s interval, the next country's name appeared on the screen, and the cycle was repeated. Unique random orderings of the 99 test countries were created for each subject for the knowledge rating and initial and final population estimation tasks.

The learning task consisted of four study-test blocks. In the study phase of each block, subjects were shown the population of each of the 24 seed countries and were given 6 s to study each one. Then, in the test phase of the block, each country's name was displayed, and subjects were to recall each country's population as exactly as possible. The order of presentation was randomized separately for each study block; the order within the test blocks was also random, with the exception that countries that appeared in the last three positions in a study block could not appear in the first three positions of the immediately following test block. Beyond the implicit information provided by seeing the seed facts in the learning phase, subjects received no feedback on the accuracy of their responses.

Subjects. Twenty-five subjects took part in Experiment 2. One was eliminated because his data suggested that he began to respond at random during the second estimation task. All subjects were Carnegie Mellon University undergraduates who participated for course credit. Subjects were run individually in sessions that lasted about 1 hr. The experiment was conducted in December 1989.

Results and Discussion

Pretest. As shown in Tables 1 and 2, pretest population estimates, both over all countries and for individual countries, closely paralleled those in Experiment 1. The rank-order correlation between median estimates of each country's population in Experiments 1 and 2 was .86 ($p < .0001$). The rank-order correlation between mean knowledge rating in the two experiments was .97 ($p < .0001$). Also as in Experiment 1, few estimates were very close to the correct values. Fewer than 6% were within 10% of the correct value.

Not surprisingly, given these results, the same factors predicted the magnitude of the population estimates in the two experiments. As shown in Table 3, the same three predictors that accounted for 73% of the variance in population estimates in Experiment 1 accounted for 72% of the variance in Experiment 2. In each case, rated knowledge (our measure of availability) accounted for the greatest amount of unique variance, true area for the next most, and true population for the least.

Learning. Exposure to the seed facts produced considerable learning. For the 24 seed set countries, the percentage of estimates within 10% of the country's true population increased from 6% on the pretest to 37%, 52%, 64%, and 74% on the four test blocks during the learning phase. This learning provided the necessary base for testing the hypothesis that learning seed facts would lead to improved understanding of metric and mapping properties for the transfer set.

Pretest–posttest changes in metric properties. Seeding the knowledge base was expected to improve both metric and mapping properties of estimates for the transfer countries (the 75 countries that were not part of the seed set). The results consistently supported the first prediction but not the second: Metric properties improved, but mappings did not.

The MOD data (computed over the 75 transfer countries) were one source of evidence for improved metric knowledge. From pretest to posttest, MOD decreased 69% (18.8 million to 5.9 million), a significant improvement, $t(23) = 2.92$, $p < .01$.

Data on individual change patterns closely linked the improved metric knowledge to subjects revising their distributional assumptions in response to experience with the seed set. Our basic hypothesis was that people would learn from the seed facts whether their initial estimates were too high or too low. Subjects whose initial estimates of the seed countries' populations were too low were expected to raise their estimates for the transfer countries; those whose initial estimates for the seed countries were too high were expected to lower their estimates for the transfer countries. Subjects also could learn whether the variability of their initial estimates for the seed countries was too high or too low and could make the appropriate adjustment for the transfer countries. The basic implication was that for all of these metric properties, differences between initial estimates and true populations for the seed countries would be strongly, and negatively, correlated with changes from initial to final estimates for the transfer countries.

This prediction proved to be correct. Changes in individual subjects' mean estimates from pretest to posttest for the transfer countries correlated $r = -.90$ with differences between their initial estimates and true values for the seed countries. The corresponding correlation for changes in the median was $r = -.87$; that for the standard deviations was $r = -.76$, indicating that subjects also learned whether their initial estimates were too variable or not variable enough. The relation for changes in the median is illustrated in Figure 4.

Because the seed set was quite representative of the full set of countries, it should not be surprising that correlations between initial estimates for the transfer countries and changes between

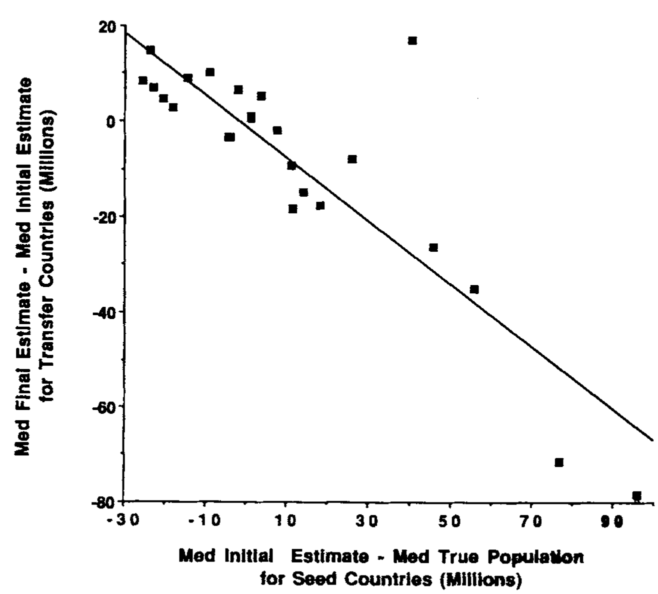

Figure 4. Relation between amount of original overestimate (underestimate) for seed countries and subsequent revision of estimate for transfer countries in Experiment 2. (Med = median.)

pretest and posttest estimates also were very high: $r = -.98$ for both the mean and the median, and $r = -.76$ for the standard deviation.

Pretest–posttest changes in mapping properties. Contrary to expectation, there was little improvement in subjects' mapping of populations onto the appropriate countries. Before subjects encountered the seed set, the median rank-order correlation between true and estimated populations for the transfer countries was .41 (ranging from .17 to .65). After subjects encountered the seed set, the corresponding median was .43 (ranging from .12 to .58). The difference between pretest and posttest correlations was not significant.

The influence of availability had been expected to decrease with the predicted increase in correlations between true and estimated populations. The fact that the correlation between true and estimated population did not increase undermined the basis of the prediction. The median of the 24 subjects' correlations between availability and estimated population was actually slightly higher after exposure to the seed countries than it had been before (.55 vs. .52).

The results also illustrated the limits of the mean absolute error measure. The 24 subjects' mean absolute error for the transfer countries declined substantially following exposure to the seed populations: from 20.9 million to 10.9 million, a significant difference, $t(23) = 2.45$, $p < .05$. If this measure had been the only one, we would have concluded that presentation of the seed countries had had the anticipated beneficial effect. However, the combination of MOD and rank-order correlation measures revealed that the reduced absolute error was entirely due to learning of metric properties; subjects did not improve their understanding of which countries went with which populations. These results again demonstrated the usefulness of the conceptual distinction between metric and mapping components of estimation.

Experiment 3

Why did exposure to the seed facts in Experiment 2 not influence the mappings of countries onto populations? One possibility was that the seed facts did not suggest clear generalizations that could be projected onto the transfer countries. Seed countries were selected to represent different levels of initial accuracy and knowledge, rather than because they suggested particular conclusions about how estimates should be revised. Encountering such a set was sufficient for subjects to abstract metric properties of the distribution. However, mappings may improve only when seed facts are selected to convey a clear and consistent lesson that disconfirms subjects' misconceptions and suggests clear alternatives.

The results of Experiments 1 and 2 indicated that one common bias is to overestimate populations of small European countries, such as Switzerland and Sweden, and to underestimate populations of East Asian countries, such as Vietnam and Thailand (see Table 1). Providing evidence inconsistent with this regional bias seemed likely to improve the mapping between those countries and their populations; providing evidence consistent with the bias seemed likely to worsen it. Specifically, exposure to disconfirming population facts (i.e., small European populations and large East Asian ones) would improve the rank-order correlation between estimated and true population for transfer countries. Exposure to bias-confirming population facts (i.e., large European populations and small East Asian ones) would have the opposite effect.

If encountering seed countries led to revised regional generalizations, these effects would be apparent on transfer countries within the seeded regions of Asia and Europe. However, there was no clear reason to expect that encountering seed facts would lead to revised mappings for transfer countries in Africa and Latin America (here labeled the *unseeded regions*, because the seed facts were not drawn from them).

Together, these expectations led to a very specific prediction: a Seed Set × Transfer Region × Trial Block interaction for correlations between true and estimated populations. Over trial blocks, correlations for transfer countries in the seeded regions should increase in the disconfirming facts condition, decrease in the confirming facts condition, and remain constant in the control condition. In contrast, for transfer countries in the unseeded regions, correlations should remain constant over trial blocks regardless of experimental condition.

We also expected the seeding to affect metric properties of the estimates. If such learning occurred, it would be evident in a decrease in MOD over trial blocks for subjects in both the bias-confirming and bias-disconfirming conditions, and there would be no similar improvement for subjects in the control condition (who did not see any seed facts). This pattern would be expected to hold for both the seeded and unseeded regions if, as the Experiment 2 results suggested, subjects' generalizations about metric properties applied to populations in general. Thus, for the MOD measure, we predicted that there would be a Seed Set × Trial Block interaction and that there would not be a three-way interaction among seed set, trial block, and transfer region.

Method

Design. Experiment 3 involved a 3 (seed set: bias-disconfirming, bias-confirming, or control) × 2 (transfer region: seeded or unseeded) × 4 (trial block: 1–4) factorial design. Type of seed set was a between-subjects variable. Subjects who received the bias-disconfirming seed set encountered three lightly populated, relatively well-known, European nations (e.g., Switzerland) and three densely populated, relatively obscure, East Asian ones (e.g., Thailand). Subjects who received the bias-confirming seed set encountered three heavily populated, well-known, European countries (e.g., West Germany) and three lightly populated, little-known, East Asian countries (e.g., Cambodia). Finally, subjects in the control condition were not shown any seed facts (Table 5).

Trial block and transfer region were within-subjects factors. On each of the four trial blocks, subjects estimated the populations of 36 countries. These were the 12 transfer countries from the seeded regions of Europe and East Asia, 12 transfer countries from the unseeded regions of Africa and Latin America, and the 12 seed countries.

Procedure. The procedure began much like that of Experiment 2. First, subjects performed the standard knowledge rating task for the 99 countries. Then, in Trial Block 1, they provided initial population estimates for the 36 countries in Table 5.

Thereafter, the Experiment 3 procedure diverged from that used previously. At the beginning of Block 2, subjects in the bias-disconfirming and bias-confirming groups were shown two seed facts. Those in the bias-disconfirming group saw the names and populations of the

Table 5
Experiment 3 Seed Sets and Transfer Countries

Country	Population (millions)
Bias-disconfirming seed set	
Switzerland	6.6
Sweden	8.2
Netherlands	14.7
Vietnam	65.2
Philippines	63.2
Thailand	54.7
Bias-confirming seed set	
West Germany	61.2
Italy	57.3
Great Britain	57.1
Cambodia	6.7
Sri Lanka	16.6
Malaysia	17.0
Transfer countries: Seeded regions	
Norway	4.2
Denmark	5.1
Austria	7.6
Belgium	9.9
Greece	10.1
Portugal	10.3
Burma	41.1
South Korea	42.6
Pakistan	107.7
Bangladesh	109.5
Japan	122.7
Indonesia	177.4
Transfer countries: Unseeded regions	
Honduras	4.8
Chad	4.8
Bolivia	6.9
Zimbabwe	9.7
Ecuador	10.2
Ivory Coast	11.2
Argentina	32.5
South Africa	35.1
Ethiopia	48.3
Mexico	83.5
Nigeria	111.9
Brazil	144.4

single least populated European and the single most populated Asian seed country. Those in the bias-confirming group saw the names and populations of the least populated Asian and the most populated European seed country. At the beginning of Block 3, subjects in these conditions saw listed above the initial seed facts the names and populations of these two countries and of the next most and next least populated country in the condition. At the beginning of Block 4, these subjects saw all six of the seed countries for their condition. Once seed facts were presented, they remained visible for the rest of the experiment. Subjects in the control condition were presented the identical four-block estimation task, except that they were not presented any seed facts.

The presentation order of the test countries was randomized for each subject and each trial block. There was a 1-min break between blocks. Beyond the seed facts, no feedback on the accuracy of estimates was provided.

Subjects. Sixty subjects, 20 in each experimental condition, participated. They were randomly assigned to condition and run individually. All were Carnegie Mellon University undergraduates. Some participated for course credit; others were paid to participate. Sessions lasted approximately 45 min. The experiment was conducted in March and April of 1990.

Results and Discussion

Metric properties. We predicted that MOD would improve for transfer countries in both seeded and unseeded regions. To test this prediction, the MOD data were submitted to a Seed Set × Transfer Region × Trial Block analysis of variance (ANOVA). The predicted Seed Set × Trial Block interaction for the MOD measure was present, $F(6, 171) = 3.37$, $p < .01$. Equally important, the Seed Set × Trial Block × Transfer Region interaction was not significant, $F(6, 171) < 1$. As shown in Figure 5, MOD for the transfer countries decreased sharply when subjects were exposed to seed facts, regardless of whether the transfer countries were drawn from the seeded regions (Europe and Asia) or the unseeded regions (Africa and Latin America). It also decreased after exposure to either bias-confirming or bias-disconfirming seed populations. Averaging over the two groups that received seed facts, mean MOD declined from 28.5 million in Block 1 to 9.6 million in Block 4, a reduction of 66%. In contrast, in the control condition, mean MOD increased 25%, from 21.6 million in Block 1 to 27.1 million in Block 4.

At the level of individual change patterns, the results replicated those of Experiment 2. As previously, the difference between each subject's pretest median and the true median for the seed countries was extremely highly correlated with changes from Block 1 to Block 4 in the subject's median estimate for the transfer countries: $r = -.90$ for the bias-disconfirming group, and $r = -.96$ for the bias-confirming group. Similarly, for subjects encountering the seed facts, median Block 1 estimate for the transfer countries correlated strongly and negatively with changes between Blocks 1 and 4 in the median estimate ($r = -.98$ and $r = -.97$ for the bias-disconfirming and bias-confirming groups, respectively).

No similar change occurred for subjects in the control group, however. Their median Block 1 estimate for the transfer countries correlated only weakly, and in the opposite direction, with changes from Block 1 to Block 4 in their median estimate ($r = .25$). The difference between this weak positive correlation and the strong negative ones for the two groups that encountered seed facts essentially rules out the possibility that regression to the mean accounted for the high correlations here or in Experiment 2. If such regression explained the pattern of changes, why would it not be evident in the changes of subjects in the control group? Instead, the difference appeared to be due to subjects observing the discrepancies between their estimates and the true populations of seed countries and adjusting their estimates for transfer countries accordingly.

Mapping properties. To test the predicted Seed Set × Transfer Region × Trial Block interaction for the rank-order correlations, we computed eight rank-order correlations between true and estimated population for each subject (separate correlations for seeded and unseeded regions for each of the four trial blocks). The resulting correlations were transformed

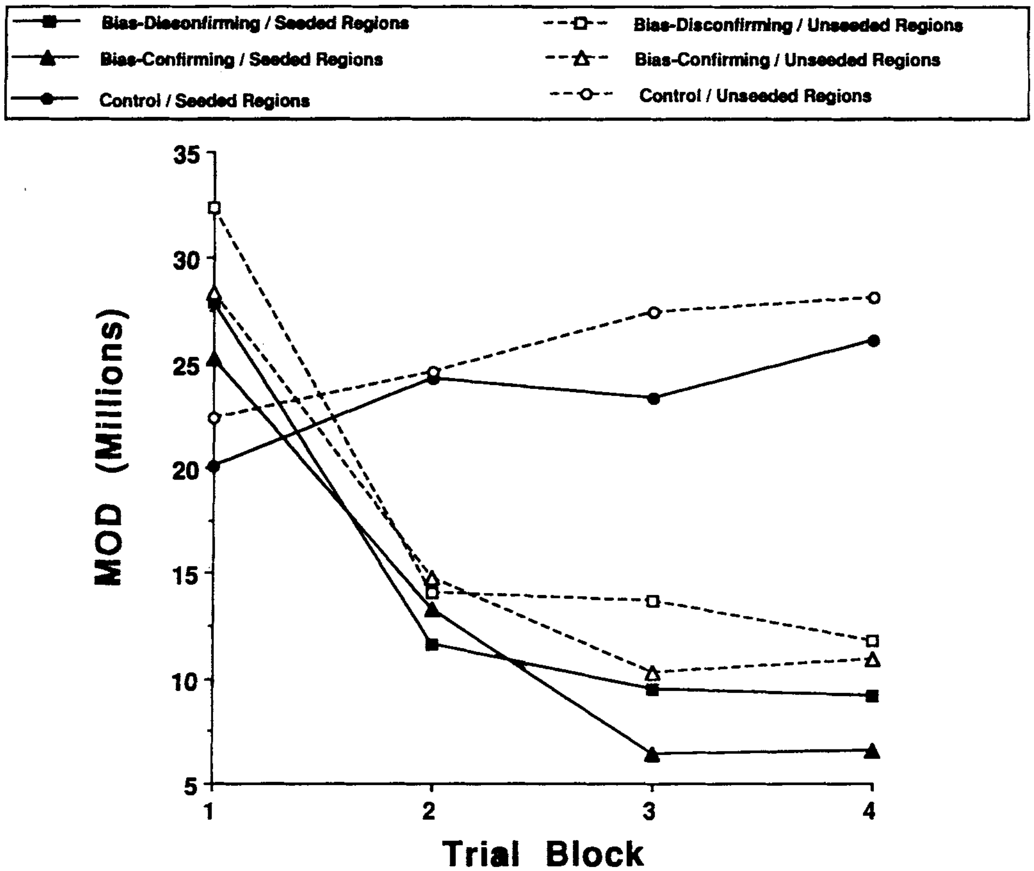

Figure 5. Change in median overall deviation (MOD) over trial blocks in Experiment 3.

using Fisher's r-to-z method and analyzed with an ANOVA (Rosenthal & Rosnow, 1991).

The predicted three-way interaction among seed set, transfer region, and trial block emerged, $F(6, 171) = 4.39$, $p < .001$, and in the anticipated form (Figure 6). Rank-order correlations for transfer countries in seeded regions increased when subjects were presented with bias-disconfirming seed populations, decreased when they were presented with bias-confirming seed populations, and stayed constant in the control condition. Also as predicted, rank-order correlations for countries from the unseeded regions of Latin America and Africa were unaffected by the presence of either confirming or disconfirming facts involving European and Asian populations.

These results provide additional evidence for the usefulness of the distinction between metric and mapping components of estimation. MOD, a measure of metric knowledge, improved under (a) conditions that led to better mappings between true and estimated population (i.e., estimates for transfer countries in seeded regions following exposure to bias-disconfirming seed facts), (b) conditions that led to worse mappings (i.e., estimates of populations of transfer countries in seeded regions following exposure to bias-confirming seed facts), and (c) conditions that led to unchanged mappings (i.e., estimates of populations of transfer countries in unseeded regions after encountering either type of seed fact). Clearly, understandings of metric and mapping properties respond differently to the same experience.

Regional generalization seemed to play a strong role in people's translation of the seed facts into mapping knowledge. Subjects generalized data on European and East Asian seed countries onto European and East Asian transfer countries, but not onto African and Latin American ones. This was the hypothesis that motivated the experiment, and the data were consistent with it.

Experiment 4

Results of Experiment 3 indicated that encountering small sets of consistent and informative population facts can substantially influence mappings of population estimates onto countries. They also suggested that the influence was due to revised regional generalizations. This raised the question of whether simply stating the regional generalization would not be at least as helpful for improving mappings as providing population facts. After all, if the purpose of carefully selecting the population facts is to allow a clear conclusion, why not simply state it?

In Experiment 4, we attempted to answer this question by examining the separate and combined effects of encountering numerical data and general statements. On the basis of the results of Experiments 3, we expected that providing carefully

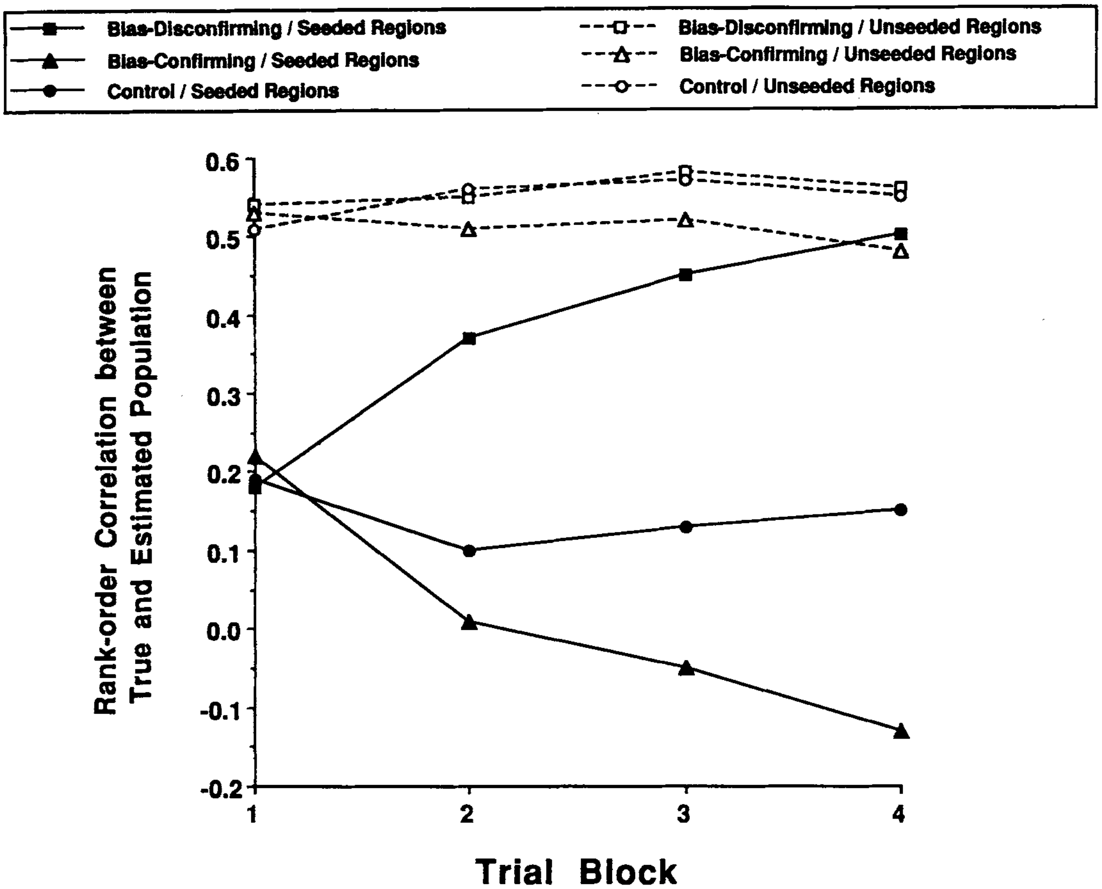

Figure 6. Change over trial blocks in mean correlations between true and estimated population in Experiment 3.

selected numerical data would influence estimates of transfer countries' populations, both by improving mappings of countries onto populations and by improving knowledge of metric properties of the distribution. Results of Experiment 3 also suggested that exposure to the numerical data would influence MOD but not rank-order correlations for transfer countries in the unseeded regions (Africa and Latin America).

How, then, would general statements, such as "European populations are usually overestimated" and "Asian populations are generally underestimated," influence estimates? On the one hand, such statements seemed likely to increase correlations between true and estimated populations for a transfer set in which Asian populations were larger than European ones (the transfer set used in Experiment 4). On the other hand, there was nothing in the general statements that would lead subjects to change metric properties of their estimates. Thus, we predicted that the general statements would not influence metric properties such as MOD for transfer countries in either seeded or unseeded regions. There also seemed little reason to expect that providing the general statements would influence rank-order correlations for countries in the unseeded regions (because no clear mapping was evident between the general statements about European and Asian population estimates and the populations of Latin American and African countries).

The factorial design used in Experiment 4 allowed us to examine the combined effect of encountering numerical data and general statements, as well as their separate contributions. It seemed likely that their combined effect on the correlations between true and estimated populations for transfer countries in seeded regions would be greater than either effect alone. General statements make explicit what is only implicit in seed facts; they guarantee that all subjects will become aware of the tendency to overestimate European and underestimate Asian populations. However, encountering seed facts may have an even greater advantage: They indicate the relevance of the general pattern to the individual subject. Simply being told that most people overestimate European populations and underestimate Asian ones has uncertain implications for any given individual. One uncertainty is whether the general statement applies; perhaps he or she is an exception. Even if the statement does apply, its implications are uncertain; how big a revision is called for? In contrast, when subjects see that the estimates they generated (or would generate) are twice as large for European countries and half as large for Asian countries as they should be, the applicability of the lesson is clear and the amount of adjustment required also becomes evident. Thus, even regarding the correlations between true and estimated populations in seeded regions, where both manipulations were expected to help, the numerical data were expected to convey benefits beyond those of the general statements.

Method

Design. Experiment 4 involved a 2 (seed populations: present or absent) × 2 (general statements: present or absent) × 2 (transfer region: seeded or unseeded) × 4 (trial block: 1–4) factorial design. The first two factors were between-subjects variables; the last two were within-subjects variables. Subjects were randomly assigned to the four conditions created by the combinations of the two between-subjects factors. Those who received seed populations saw the names and populations of three lightly populated European countries (Switzerland, Sweden, and the Netherlands) and three heavily populated Asian countries (Thailand, Vietnam, and Bangladesh). Those who received the general statements were told "European populations tend to be overestimated" and "Asian populations tend to be underestimated." Population estimates were obtained on each of four trial blocks for each of 36 countries. They were the same countries examined in Experiment 3 and listed in Table 5.

Procedure. The Experiment 4 procedure was almost identical to that used in Experiment 3. Subjects rated their knowledge of the 98 "4 million plus countries." (This experiment followed the reunification of Germany, which reduced the number of countries with more than 4 million people by one.) Then they were presented the four trial blocks. During Block 1, subjects in all conditions provided initial estimates of the population of each country. At the beginning of Block 2, subjects in the general-statements conditions were told about the typical biases in estimates of European and Asian populations. Also at the beginning of Block 2, subjects in the seed-populations conditions were presented with all six seed countries' populations. Subjects who received both general statements and seed facts encountered both types of information on the same display. Once seed populations, general statements, or both were presented, they remained visible for the duration of the experiment.

Subjects. Eighty Carnegie Mellon University undergraduates participated in this experiment, some for course credit and others for pay. Equal numbers were randomly assigned to each of the four experimental groups. The experiment was conducted in March and April of 1991.

Results and Discussion

Metric properties. A 2 (seed populations: present or absent) × 2 (general statements: present or absent) × 2 (transfer region: seeded or unseeded) × 4 (trial block) ANOVA was conducted on the MOD data. As predicted, the analysis showed a significant positive effect on this measure of metric knowledge for exposure to the seed populations, $F(1, 76) = 5.49$, $p < .05$, and not for exposure to the general statements ($F < 1$). MOD also improved over the four trial blocks, $F(3, 228) = 5.22$, $p < .01$, with the large majority of the improvement coming between the first and second blocks (when the seed populations, general statements, or both were provided). A significant Trial Block × Seed Populations interaction was also present, $F(3, 228) = 5.20$, $p < .01$. On the first trial block, before receiving the population facts, the four groups did not differ; on all subsequent trials, they did. The MODs of subjects who received the seed facts decreased 51% between Trial Blocks 1 and 4; mean values of MOD on the four blocks were 24.6 million, 12.7 million, 11.9 million, and 11.9 million, respectively. The comparable figures for subjects who did not receive seed facts were 25.8 million, 24.9 million, 26.8 million, and 25.5 million, respectively.

As in the two previous experiments, examination of individual subjects' performance closely linked changes in median estimates for the transfer countries to their learning from the seed facts. In the group exposed to both seed populations and general statements, differences between initial estimates and true populations for the seed countries correlated $r = -.90$ with changes over the four trial blocks in median estimates for transfer countries. The corresponding correlation for the group that received only the seed facts was $r = -.78$. Correlations between median Block 1 estimates for the transfer countries and changes between Blocks 1 and 4 also were very high: $r = -.93$ for the group that was presented seed populations and general statements, and $r = -.84$ for the group that was presented only the seed populations. The corresponding correlation for the control group was $r = -.01$.

Mapping properties. A Seed Populations × General Statements × Transfer Region × Trial Block ANOVA was conducted on the rank-order correlations between true and estimated populations (r-to-z transformed data). The analysis indicated a pair of significant three-way interactions: a Seed Populations × Transfer Region × Trial Block interaction, $F(3, 228) = 4.91$, and a General Statements × Transfer Region × Trial Block interaction, $F(3, 228) = 5.89$.

Figure 7 illustrates the sources of these interactions. As indicated in the top panel, the numerical data and the general statements both led to increasing correlations between true and estimated populations over trial blocks for transfer countries in the seeded regions. The two variables combined in an approximately additive fashion, as is apparent both in the figure and in the nonsignificance of all interactions involving both variables. In contrast, but also as expected, correlations between true and estimated populations for transfer countries in the unseeded regions did not change over trial blocks in response to either the seed populations or the general statements (lower panel, Figure 7).

These results provide additional evidence for the distinction between metric and mapping components. There were conditions in which both metric and mapping properties of the estimates improved (i.e., responses to countries in the seeded regions by subjects who received numerical population data), conditions in which metric properties improved but mapping properties did not (responses to countries in the unseeded regions by subjects who received numerical population data), and conditions in which mapping properties improved but metric ones did not (responses to countries in the seeded regions by subjects who received only the general statements). Knowledge of both metric and mapping properties is essential for accurate estimation, but the two types of knowledge clearly are distinct, both conceptually and empirically.

Experiment 4 differed from Experiment 3 in that subjects were exposed to all six seed populations at once, rather than encountering two, then four, then all six. This indicates that the previously observed effects of encountering seed populations were not limited to the sequential presentation procedure used in the earlier experiments. It also indicates that the gradual rise over trial blocks in correlations between true and estimated populations that was observed in Experiment 3 was due to the number of available seed facts rather than to the implications of the pattern being realized increasingly over trial blocks. When

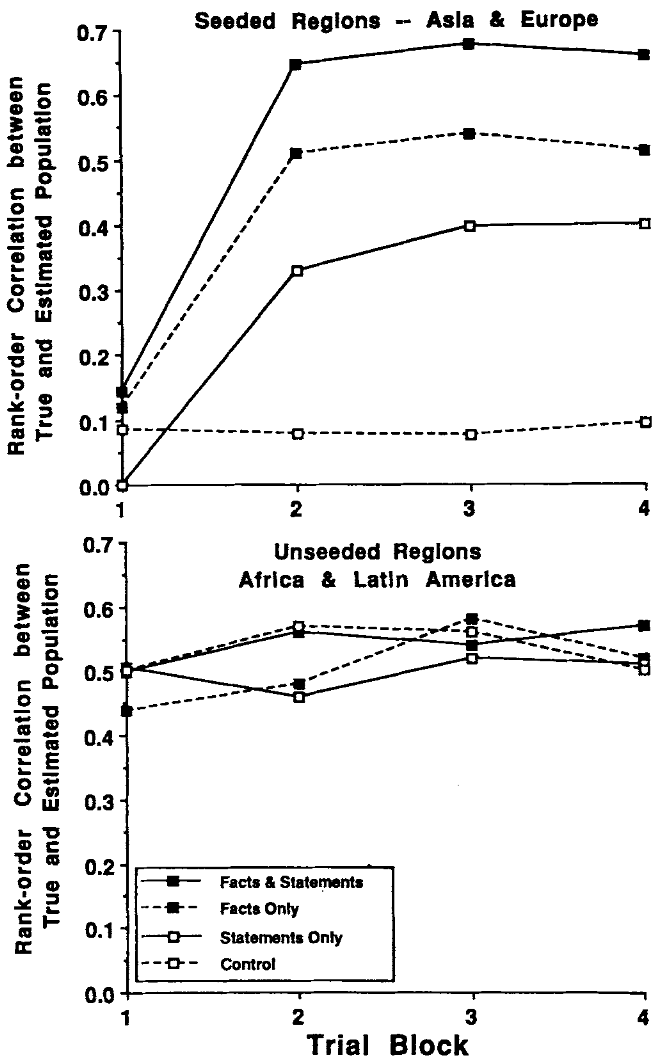

Figure 7. Change over trial blocks in mean correlations between true and estimated population in Experiment 4.

all quantitative information was provided before Trial Block 2, both mapping and metric properties improved greatly, and little if any further improvement was seen subsequently on either measure.

Finally, the results of Experiment 4 indicate that exposure to relevant quantitative data confers benefits beyond those obtained by simply being exposed to general statements that summarize the basic pattern. Quantitative data indicate the relevance of generalizations to the individual subject, the amount of revision that is needed in future estimates, and the right range for estimates in the particular domain. For these reasons, exposure to carefully selected quantitative facts may be especially useful for helping people improve their estimates.

General Discussion

Beyond the specific findings that emerged in the present experiments, we see the research as making three general contributions. First, the metrics and mappings framework provides a means for analyzing real-world quantitative estimation that should prove applicable in numerous domains. Second, the cue-validity portion of the framework provides a means for integrating the influences of heuristics and domain-specific knowledge and for predicting the contexts in which each will be most influential. Third, seeding the knowledge base provides an instructional technique for improving estimation that again should be applicable in a broad range of real-world contexts. We discuss each of these contributions in the following paragraphs.

Metrics and Mappings Framework

Previous research on real-world quantitative estimation has focused on two types of processes often used in estimation: reasoning from domain-specific knowledge and application of heuristics, such as availability, representativeness, and anchoring. This past research has demonstrated that each type of process influences estimation. However, the past research has told us little about the conditions under which each process is most influential, about how the processes work together to produce estimates, or about how best to conceptualize estimation as a whole.

The metrics and mappings framework goes beyond this previous work by providing an integrative perspective for thinking about these and other influences on real-world quantitative estimation. Its basic assumption is that estimates are derived from two types of knowledge: knowledge of the distribution of quantitative values in the domain and knowledge of the relative positions of particular entities within that distribution.

This simple distinction proves to have surprising power for analyses of estimation. Metrics and mappings are independent, both conceptually and empirically. Some types of information lead to improvements in both, some improve one but not the other, and some improve neither. The framework allows straightforward prediction of which experiences will influence each component. Experiences that convey distributional information—information about central tendency, variability, range, or shape of a distribution—should advance metric knowledge. Such information presumably can be conveyed through direct description of the distribution (as in a scatterplot), through descriptions of features of the distribution (as in being told the mean or range), or through induction from particular quantitative examples (as in inducing the shape of the distribution from being told or shown numerical values in it). In contrast, mapping knowledge is influenced by experiences that convey information about relative status of individual entities (as in being told that the Philippines has a larger population than France), through statements concerning categorical relations (as in being told that Asian countries tend to have more people than European countries), or through induction of those same categorical relations from specific examples.

A particularly important aspect of the metrics and mappings framework is its generality. It can be used to analyze any domain in which people estimate quantities. The examples used in this article—amount of television watching by different age groups, baseball batting averages, national debt statistics, populations, and land areas—provide some sense of the variety of these domains. The following list includes a few of the other quantitative dimensions and content areas in which estimation could be studied within the same framework: distances between different cities, temperatures of different classes of stars, dollar sales of different products, frequencies of different sounds, densities of different elements, caloric intakes in different societies, running speeds of different animals, and numbers of immigrants entering the United States in different years. We recently have made a first step toward demonstrating this generality by showing the applicability of the framework to students' estimates of distances between cities (Brown & Siegler, 1993).

The distinction between metric and mapping information also raises a number of issues that the present research only begins to address. One concerns the exact distributional knowledge that people acquire from being exposed to metric information. Thinking in terms of distributions makes possible a systematic approach to the types of knowledge people acquire through intuitive statistical induction. It raises such issues as whether people are sensitive to differences among alternative measures of central tendencies (means, medians, and modes), whether they are sensitive to different moments of the distribution (mean, variance, skewness, and kurtosis), and whether they differentiate among general shapes of distributions (normal, rectangular, bimodal, and so forth). The distinction also raises the issue of whether people represent their knowledge of the quantitative distribution separately from their knowledge of specific entities within it or whether the distributional knowledge just falls out from beliefs about the individual entities. Knowing the answers to these questions would considerably deepen our understanding of real-world quantitative estimation.

Cue Validities and Estimation

Within the metrics and mappings framework, mapping knowledge is hypothesized to be derived in large part from a learning of the relative validities of alternative cues. Specifically, it is assumed that people typically use multiple cues to derive an estimate, that they weigh some of these cues more heavily than others, and that the weight they assign to each cue is a function of its predictive accuracy relative to the predictive accuracy of competing cues.

Applying cue-validity concepts to real-world quantitative estimation leads to the recognition that both heuristics and domain-specific knowledge often influence the value of a single estimate. They operate as alternative cues, rather than as qualitatively different types of information. In Experiment 1 of the present study, the notion that cue utilization is determined in part by ecological validity allowed us to predict that the availability heuristic would be emphasized in population estimates, that specific knowledge of maps would be emphasized in area estimates, and that both types of information would be used to some extent in both situations. These findings support J. R. Anderson's (1990) theoretical arguments that analyses of the structure of the environment can help predict people's relative weighting of cues.

The technique of assessing ecological validities to predict such weightings seems applicable to many domains and heuristics. It can be used to predict when a given heuristic is likely to be used, how heavily it will be weighed, and how it will bias performance. For example, the representativeness heuristic (Kahneman & Tversky, 1972; Tversky & Kahneman, 1974) seems amenable to an analysis like that provided in this article for availability. Such an analysis follows from recognizing that representativeness often reduces to typicality (Collins & Michalski, 1989; Shafir, Smith, & Osherson, 1990). In the same way that knowledge ratings provide an index of an item's availability, typicality ratings would provide an index of its representativeness. This would allow computation of the predictive validity of representativeness for a given quantitative dimension defined over a class of items. Specifically, the predictive validity of typicality in a given domain would be the correlation between each item's typicality rating and that item's true quantitative value on the dimension of interest. That correlation, along with the correlations involving competing cues, could then be used to predict the presence and strength of representativeness effects in that domain.

The present framework implies that representativeness, like availability, should play an important role when it is strongly predictive of the to-be-estimated quantity relative to alternative cues; it should play a minor role when it is weakly predictive relative to the alternatives. Large representativeness biases would be predicted when representativeness is the most predictive of the available cues but is still, in an absolute sense, not extremely predictive of the to-be-estimated quantity. For example, the cues that are stereotypically associated with being an engineer may be only weakly predictive of whether an individual is an engineer or a lawyer, but no better personality-based cue may be present. This, together with a lack of experience in using base rates to predict individual characteristics, may be what leads to reliance on representativeness in the engineer-or-lawyer judgment task (Kahneman & Tversky, 1972).

Anchoring-and-adjustment, the third of Tversky and Kahneman's (1974) classic heuristics, can be analyzed similarly. Anchors are domain-specific quantitative facts (e.g., "The population of Switzerland is 6.6 million"). When such a fact is retrieved in the course of generating an estimate, its weighting is likely to depend on the similarity between the target item and the anchor item (or items) along the relevant dimensions and on the similarity and predictive value of competing cues. When target and anchor are very similar (e.g., Switzerland and Austria), the anchor is likely to be weighed heavily in generating estimates. In contrast, the anchor is likely to be weighed less when target and anchor are dissimilar (e.g., Switzerland and Kenya).

The differential impact of the seed sets on estimates of populations of transfer countries in seeded and unseeded regions (Experiments 3 and 4) illustrates this principle at work. It seems reasonable to assume that subjects considered the seed countries (which were by definition from the seeded regions) to be at least fairly similar to the transfer countries in the seeded regions. It also seems reasonable to assume that they saw little resemblance between seed countries and those transfer countries that were from unseeded regions. Consistent with these assumptions, correlations between true and estimated populations of transfer countries from the seeded regions were greatly affected by exposure to the seed sets, but correlations for countries from unseeded regions were not. These results suggest that subjects relied heavily on the seed facts in revising their relative estimates for the (similar) European and Asian countries from the seeded regions and that they assigned little if any weight to them in revising relative estimates for the (dissimilar) countries from the unseeded regions of Africa and Latin America. Obtaining similarity ratings along the relevant dimension for seed and transfer countries might facilitate prediction of the amount and direction of revision of estimates for each transfer country as a function of encountering specific seed facts.

In summary, the three most prominent judgmental heuristics—availability, representativeness, and anchoring-and-adjustment—can all be analyzed straightforwardly within a cue-validity framework. The predictive strength of availability for a particular estimation task can be measured by correlating rated knowledge of each entity in the domain with the quantitative values of those entities on the relevant dimension (as illustrated in the present study). The predictive strength of representativeness can be measured by correlating each item's rated typicality with that item's quantitative value on the relevant dimension. The predictive strength of sets of anchor values can be measured by computing their rated similarity to target items along the relevant dimension. In principle, knowing the predictive strengths for these cues and for competing domain-specific cues should allow accurate prediction of the relative weights given to each type of information.

Educational Implications

Acquisition of quantitative knowledge poses a substantial educational challenge. Many of the facts students need to know in the physical, biological, and social sciences are quantitative: How far is the moon from the earth? What is the atomic weight of a mole of sulphuric acid? When did the Paleolithic Period end? When was the Magna Carta signed? What were the populations of the Union and the Confederacy at the beginning of the Civil War? This quantitative information is far from trivial. Without it, many issues are impossible to understand. For example, the military superiority of the North, despite early uninspired leadership and the continuing disadvantage of fighting far from home, is difficult to understand without comprehending the extent of differences in populations and industrial capacity of the North and South.

Faced with the issue of how to inculcate such information, educators have oscillated between two approaches. One has been to require students to memorize large numbers of quantitative facts. The other has been to deemphasize dates, magnitudes, and other quantities and to focus on understanding of qualitative relations. Each of these approaches has major drawbacks, however. On the one hand, the motivational and cognitive problems associated with memorization of large numbers of facts, quantitative or otherwise, are large and probably unavoidable. Also, there are just too many such facts for anyone to memorize a high percentage of them. On the other hand, it is difficult if not impossible to acquire more than a superficial understanding of a domain without some degree of quantitative sophistication about it.

The present research points to an alternative: the *key facts approach.* The central tenet of this approach is that students need to learn some quantitative facts, but not a vast number of them. Instead, it should be possible to promote a good degree of quantitative understanding by exposing students to relatively small, carefully selected, sets of such facts. Learning the quantitative values of these key facts would serve students well in several ways. First, the quantities would provide the data necessary for inducing metric properties of the domain. Second, they would invite revision of prior beliefs and generation of new beliefs concerning factors that underlie the relative quantities. Third, they would serve as explicit quantitative reference points, useful for drawing general lessons about values of unstudied items. Finally, presenting such facts and not others should focus attention on the most important and representative information, and hence avoid interference from less important and representative facts.

Of course, these benefits will not come for free. Selection of key facts in a domain requires an understanding of the factors that determine quantitative values in the domain, an understanding of students' prior knowledge and beliefs about these determining factors, and identification of particular information that will disconfirm prior misconceptions and move beliefs in the right direction. These are not trivial objectives. Results of the present research, however, suggest that the effort is worth making; the reward can be dramatic improvement in students' ability to estimate quantitative features of the world in which they live.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, N. H. (1964). Test of a model for number-averaging behavior. *Psychonomic Science, 1,* 191–192.

Baddeley, A. D., Lewis, V., & Nimmo-Smith, I. (1978). When did you last. . .? In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 77–83). San Diego, CA: Academic Press.

Beach, L. R., & Scopp, T. S. (1968). Intuitive statistical inferences about variances. *Organizational Behavior and Human Performance, 3,* 109–123.

Beach, L. R., & Swenson, R. G. (1966). Intuitive estimation of means. *Psychonomic Science, 5,* 161–162.

Brown, N. R. (1990). The organization of public events in long-term memory. *Journal of Experimental Psychology: General, 119,* 297–314.

Brown, N. R., Rips, L. J., & Shevell, S. K. (1985). The subjective dates of natural events in very long-term memory. *Cognitive Psychology, 17,* 139–177.

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). *Inducing statistical properties of real-world domains.* Unpublished manuscript.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments* (2nd ed.). Berkeley: University of California Press.

Chapman, L. J., & Chapman, J. P. (1969). Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology, 74,* 271–280.

Collins, A. M. (1978a). Fragments of a theory of human plausible reasoning. In D. Waltz (Ed.), *Theoretical issues in natural language processing* (Vol. 2, pp. 194–201). Urbana: University of Illinois.

Collins, A. M. (1978b). *Human plausible reasoning* (Report No. 3810). Cambridge, MA: Bolt, Beranek and Newman.

Collins, A. M., & Michalski, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science, 13,* 1–49.

Combs, B., & Slovic, P. (1979). Causes of death: Biased newspaper coverage of biased judgments. *Journalism Quarterly, 56,* 837–843, 849.

Edwards, W. (1968). Conservatism in human information processing. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 17–52). New York: Wiley.

Ferguson, R. P., & Martin, P. (1983). Long-term temporal estimation in humans. *Perception & Psychophysics, 33,* 585–592.

Fischhoff, B. (1987). Judgment and decision making. In R. Sternberg & E. Smith (Eds.), *The psychology of human thought* (pp. 153–187). Cambridge, England: Cambridge University Press.

Friedman, W. J. (1987). A follow-up to "Scale effects in memory for time of events": The earthquake study. *Memory & Cognition, 15,* 518–520.

Friedman, W. J., & Wilkins, A. J. (1985). Scale effects in memory for time of events. *Memory & Cognition, 13,* 168–175.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528.

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics.* Hillsdale, NJ: Erlbaum.

Hammond, K. R. (1990). Functionalism and illusionism: Can integration be usefully achieved? In R. M. Hogarth (Ed.), *Insights in decision making: A tribute to Hillel J. Einhorn* (pp. 227–261). Chicago: University of Chicago.

Hendrick, C., & Costantini, A. R. (1970). Number averaging behavior: A primary effect. *Psychonomic Science, 19,* 121–122.

Hogarth, R. (1987). *Judgment and choice* (2nd ed.). New York: Wiley.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review, 98,* 352–376.

Information please almanac (42nd ed.). (1989). Boston: Houghton Mifflin.

Information please almanac (43nd ed.). (1990). Boston: Houghton Mifflin.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology, 3,* 430–454.

Kerst, S. M., & Howard, J. H., Jr. (1978). Memory psychophysics for visual area and length. *Memory & Cognition, 6,* 327–335.

Krueger, J., Rothbart, M., & Sriram, N. (1989). Category learning and change: Differences in sensitivity to information that enhances or reduces intercategory distinctions. *Journal of Personality and Social Psychology, 56,* 866–875.

Kunda, Z., & Nisbett, R. E. (1986). The psychometrics of everyday life. *Cognitive Psychology, 18,* 195–224.

Lathrop, R. G. (1967). Perceived variability. *Journal of Experimental Psychology, 23,* 498–502.

Levin, I. P. (1974). Averaging processes in ratings and choices based on numerical information. *Memory & Cognition, 2,* 786–790.

Levin, I. P. (1975). Information integration in numerical judgments and decision processes. *Journal of Experimental Psychology: General, 104,* 39–53.

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.

Linton, M. (1975). Memory for real-world events. In D. A. Norman & D. E. Rumelhart (Eds.), *Explorations in cognition* (pp. 376–404). San Francisco, CA: W. H. Freeman.

MacWhinney, B. (1987). *Mechanisms of language acquisition.* Hillsdale, NJ: Erlbaum.

MacWhinney, B., Leinbach, J., Taraban, R., & McDonald, J. (1989). Language learning: Cues or rules? *Journal of Memory and Language, 28,* 255–277.

Malmi, R. A., & Samson, D. J. (1983). Intuitive averaging of categorized numerical stimuli. *Journal of Verbal Learning and Verbal Behavior, 22,* 547–559.

McClelland, J. L., & Rumelhart, D. E. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 2. Psychological and biological models.* Cambridge, MA: MIT Press.

Means, B., & Loftus, E. F. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology, 5,* 297–318.

Means, B., Nigam, A., Zarrow, M., Loftus, E. F., & Donaldson, M. S. (1989). *Autobiographical memory for health-related events* (DHHS Publication No. PHS 89–1077). Hyattsville, MD: National Center for Health Statistics.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 32, pp. 89–115). Palo Alto, CA: Annual Reviews.

New York Times citation index (Vol. 76). (1988). New York: New York Times.

Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing* (pp. 283–308). San Diego, CA: Academic Press.

Nickerson, R. (1981). Motivated retrieval from archival memory. In J. H. Flowers (Ed.), *Nebraska Symposium on Motivation 1980* (pp. 73–119). Lincoln: University of Nebraska Press.

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin, 68,* 29–46.

Pitz, G. F., & Sachs, N. J. (1984). Judgment and decision: Theory and application. *Annual Review of Psychology, 35,* 139–163.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

Shafir, E. B., Smith, E. E., & Osherson, D. N. (1990). Typicality and reasoning fallacies. *Memory & Cognition, 18,* 229–239.

Sherman, S. J., & Corty, E. (1984). Cognitive heuristics. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 189–285). Hillsdale, NJ: Erlbaum.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology, 28,* 1–39.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1982). Facts versus fears: Understanding perceived risk. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 463–489). Cambridge, England: Cambridge University Press.

Smith, J. F., & Kida, T. (1991). Heuristics and biases: Expertise and task realism in auditing. *Psychological Bulletin, 109,* 472–489.

Spencer, J. (1961). Estimating averages. *Ergonomics, 4,* 317–328.

Spencer, J. (1963). A further study of estimating averages. *Ergonomics, 6,* 255–265.

Thompson, C. P. (1982). Memory for unique personal events: The roommate study. *Memory & Cognition, 10,* 324–333.

Thompson, C. P., Skowronski, J. J., & Lee, D. J. (1987, July). *Reconstructing the date of a personal event.* Paper presented at the Second International Conference of Practical Aspects of Memory, Swansea, Wales.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185,* 1124–1131.

Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology, 18,* 225–252.

Wallsten, T. S., & Barton, C. (1982). Processing probabilistic multidimensional information for decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 361–384.

Appendix

Table A1

True and Estimated Areas, New York Times Citation Index (1988) Counts, and Self-Knowledge Ratings for the 99 Most Populated Countries (Data From Experiment 1)

Rank	Country	Actual area (thousands of square miles)	Median estimated area (thousands of square miles)	*New York Times Citation Index*	Mean knowledge rating (0–9 scale)
1	China	3,705	3,000	470	4.71
2	India	1,269	2,000	314	4.50
3	U.S.S.R.	8,649	5,750	2,383	6.33
4	Indonesia	735	205	33	1.88
5	Brazil	3,286	1,250	188	3.46
6	Japan	144	125	807	4.88
7	Nigeria	357	188	41	1.75
8	Bangladesh	56	100	54	1.67
9	Pakistan	310	238	233	2.58
10	Mexico	762	1,200	280	5.33
11	Vietnam	127	100	152	4.08
12	Philippines	116	130	154	3.38
13	West Germany	96	500	366	4.92
14	Italy	116	300	204	5.17
15	United Kingdom	95	390	563	5.67
16	France	211	660	420	5.83
17	Thailand	198	98	60	2.00
18	Egypt	387	500	112	3.75
19	Turkey	301	450	78	2.88
20	Iran	636	425	622	3.71
21	Ethiopia	472	255	51	2.96
22	South Korea	38	110	308	3.13
23	Burma	261	117	96	0.96
24	Spain	195	550	102	4.75
25	Poland	121	201	264	3.58
26	South Africa	471	700	517	3.92
27	Zaire	906	70	23	0.96
28	Argentina	1,068	700	101	2.42
29	Colombia	440	200	105	2.58
30	Canada	3,852	4,000	379	6.67
31	Morocco	172	90	34	1.79
32	Tanzania	365	95	7	0.75
33	Algeria	920	188	58	1.21
34	Sudan	967	200	36	1.17
35	Yugoslavia	99	158	95	2.58
36	Kenya	225	211	33	2.29
37	Romania	92	108	40	1.54
38	North Korea	47	133	90	2.67
39	Peru	496	500	66	2.54
40	Taiwan	14	125	92	3.04
41	Venezuela	352	200	32	2.63
42	Nepal	54	50	17	1.29
43	Iraq	168	365	288	3.25
44	Malaysia	127	100	24	1.33
45	East Germany	42	200	75	4.17
46	Sri Lanka	25	30	53	1.58
47	Australia	2,968	1,453	87	4.46
48	Uganda	91	95	9	0.71
49	Czechoslovakia	49	188	72	2.71
50	Mozambique	309	113	34	0.75
51	Netherlands	14	100	38	2.54
52	Afghanistan	250	350	355	2.67
53	Ghana	92	100	6	0.42
54	Saudia Arabia	830	800	89	3.04
55	Chile	292	358	97	2.42
56	Syria	71	155	101	2.46
57	Ivory Coast	125	77	12	1.13
58	Madagascar	227	100	1	1.08
59	Hungary	36	245	90	2.04
60	Cameroon	184	68	2	0.63
61	Cuba	43	50	215	3.88

(*Table A1 continues on next page*)

Appendix

Table A1 (*continued*)

Rank	Country	Actual area (thousands of square miles)	Median estimated area (thousands of square miles)	*New York Times Citation Index*	Mean knowledge rating (0–9 scale)
62	Portugal	36	100	21	2.83
63	Ecuador	109	100	20	2.00
64	Greece	51	188	86	4.08
65	Belgium	12	125	35	2.63
66	Zimbabwe	151	75	36	1.04
67	Bulgaria	43	300	30	1.88
68	Guatemala	42	97	37	2.08
69	Mali	479	45	4	0.75
70	Burkina Faso	106	63	2	0.17
71	Sweden	174	495	64	3.50
72	Angola	481	140	164	0.71
73	Somalia	249	83	8	0.54
74	Tunisia	63	73	19	0.83
75	Malawi	46	86	3	0.50
76	Austria	32	325	86	3.46
77	Zambia	291	75	7	0.58
78	Niger	489	180	6	1.29
79	Rwanda	10	84	4	0.63
80	Senegal	76	60	11	0.54
81	Guinea	95	98	2	1.17
82	Bolivia	424	240	20	1.54
83	Dominican Republic	19	39	19	1.67
84	Kampuchea	70	100	85	1.58
85	North Yemen	75	60	4	0.50
86	Switzerland	16	175	51	3.71
87	Haiti	11	25	112	2.29
88	El Salvador	8	73	108	2.83
89	Burundi	11	50	12	0.17
90	Denmark	17	135	20	3.25
91	Finland	130	238	25	2.83
92	Chad	496	104	11	1.17
93	Honduras	43	98	91	2.04
94	Benin	43	35	1	0.29
95	Israel	8	95	582	4.50
96	Paraguay	157	225	18	1.79
97	Norway	125	315	40	3.38
98	Libya	679	200	52	3.04
99	Sierra Leone	28	85	1	0.29

Received March 5, 1992
Revision received October 29, 1992
Accepted November 23, 1992 ■