THE UNIVERSITY OF ALBERTA

THE ROLE OF ORDINAL INFORMATION AND THE EFFECTS OF CUE VALIDITY IN NUMERICAL ESTIMATION

BY PETER JAMES LEE ©

A THESIS SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 2005

0-494-08671-8

Your file Votre référence
ISBN:
Our file Notre référence
ISBN:

To my father, Lee Chai Seng, who did not live to see this work completed but who nonetheless encouraged and supported my academic endeavours from the beginning.

ABSTRACT

The thesis examined the role of ordinal information and the cognitive processes that underlie numerical estimation form the perspective of plausible reasoning and inference generation. Five experiments are reported here. Experiments 1 and 2 examined people's ability to identify and use a modal type of information called proxy cues, that are thought to be better known than the target variable and are highly predictive of the target domain. These two experiments demonstrated that proxy cues (such as wealth) were most often identified and frequently used to judge infant mortality rates. Nevertheless, other (often non-predictive) information was also accessed. Individual differences in the identification and use of predictive information were also identified. Experiment 3 introduced a new technique called proxy cueing. Proxy cueing was shown to influence ordinal accuracy independently from metric accuracy, and that this type of ordinal information appears to work at the level of each target item rather than evoke changes in beliefs that could affect all mortality responses. Experiment 4 showed that when ecological validity is high and target items are representatively sampled, most respondents are able to accept and use wealth information but reject non-predictive information (such as population). In contrast, when target items were sampled so that the ecologically valid information was no longer predictive, respondents appeared to employ one of two information selection modes; by either accepting ecologically valid information and rejecting population or by accepting both types of information as being internally valid for the to-be-estimated set.

These experiments show that people are often able to make use of task relevant knowledge, although there are differences in how and when they make use of predictive and non-predictive information. These differences point toward a mechanism capable of evaluating proxy information that is external to the processes used generate plausible numerical answers. To date this mechanism has not been adequately defined and suggests future research should concentrate on how information is judged to be useful, and why there are individual differences in people's ability correctly identify predictive from non-predictive information.

ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude to Norman Brown who, while on sabbatical in the U.K., took concern on my then situation by offered me his mentorship, unbridled access to his facilities and financial support, but above all his friendship.

I would also like to thank members of my committee, Alinda Friedman and Tom Spalding, and Peter Dixon. Special thanks also to the many people who have taken time to offer advice, help, and encouragement on this and many other projects, not least John Reddon, John Pracejus, and Courtney Bryden. Finally, immeasurable thanks to Carole and Roger Powell for their unstinting support before, during and beyond my studentship.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

OVERVIEW

In the following six chapters I address two issues in real-world numerical estimation that deal with how people rank order their responses, despite the fact that many of the actual values are often unknown. Subsequent to Brown and Siegler (1993) identifying ordinal accuracy as an important component in numerical estimation there have been unanswered questions about the nature of the cognitive processes involved, and the types of information people access that can be used to rank order numerical judgments (see Brown, 2002). Hence the rationale behind this project is to address topics relating to the cognitive processes involved in generating estimates under uncertainty, and the information that people use to maintain ordinal relationships between target items.

Extending our understanding about process and information is important for a number of reasons. Despite the development of a hypothesized strategy for rank ordering items, called *ordinal-conversion* (Brown, 2002), the strategy and the processes that distinguish ordinal-conversion from other estimation theories require empirical verification. Although there is a consensus that people often retrieve task relevant information, it is also an empirical but unresolved issue as to the exact nature of this information.

Understanding how people generate ordinal inferences from task relevant information also raises interesting questions about how information is evaluated, people's beliefs about the value of particular types of information, and why some types of information appear to be more relevant than others. Finally, there remains a broader interdisciplinary need to understand how ordinal judgments are made in finer detail than is currently understood. For example, ordinal judgments play an important role in the development of expert systems, data mining techniques, and decision support procedures (e.g., Moshkovich, Mechitov, & Olson, 2002). Thus, a better understanding of real-world numerical estimation may contribute towards more sophisticated decision models across both academic and applied disciplines.

In the remainder of this chapter I will provide a review of the background literature that underpins this project, and also provide a precise conceptualisation of the ordinal conversion position and explanations of the terminology used throughout this thesis. Chapter I will also introduce some hypotheses regarding one type of information that appears to be particularly useful for maintaining ordinal accuracy between to-be-estimated items. There will also be a detailed explication of an alternative position, termed *QuickEst*, which falls from the more general "*fast and frugal*" approach proposed by the ABC Research Group (Gigerenzer, Todd, & the ABC Research Group, 1999).

Chapter II temporarily puts aside questions about estimation processes by describing two experiments focusing on the type and quantity of information people use to generate quality-of-life estimates (e.g., numerical estimates for quality-of-life statistics such as life expectancy, literacy rates, and infant mortality rates in different countries). Both of these studies are exploratory. Nevertheless, they provide early evidence that despite large individual differences in the selection of task relevant information, there exists modal types of information that show characteristics that can be used to help preserve ordinal accuracy.

In Chapter III, the research aim switches to the description and testing of a new methodology for investigating ordinal-conversion called *proxy cueing*. In this Chapter, the technique is compared with an established method, termed *seeding the knowledge base,* which has typically been used to investigate changes in metric rather than ordinal accuracy (cf. Brown & Siegler, 1993). Chapters IV and V extend the use of proxy cueing by investigating how information and process combine to influence ordinal accuracy, and especially how context determines the quantity and type of information people use to make reasoned inferences. These chapters also argue against the notion that human numerical estimation, and ordinal judgments in particular, can be characterized as respondents accessing either single cues use or multiple cues alone. In turn, this argument provides grounds for questioning theories that are unable to adapt to explain differences in cue selection strategies across different contexts. Finally, Chapter VI provides an overview of the findings presented here, and addresses points of interest that were beyond the scope of the current project and outline new areas of investigation and possible future directions.

INTRODUCTION

Laboratory tasks used to investigate numerical judgments typically request participants to estimate the value for each item serially from a set of items. For example, respondents might be asked to estimate the populations of European countries, or the risk of dying from certain diseases. These tasks have an important role in understanding human inductive reasoning, because numerical values are seldom stored and hence rarely retrieved directly from memory. It follows that estimates of real-world quantities tend to rely on plausible reasoning and inference about the target items generated from task and item relevant information (Brown, 2002; Collins & Michalski, 1989; Paulos, 1990). This does not imply that accuracy is necessarily poor or that the way people generate estimates is trivial. The absence of exact knowledge requires people to recall what they know and to generate inferences, often from partial information (Collins, Warnock, Aiello, & Miller, 1975; Collins & Michalski, 1989). As a result people can often draw principled conclusions, and responses may sometimes correspond reasonably well with the actual numerical values.

Still, numerical estimates can also be wrong or biased and the true numerical values for items on a given domain may sometimes seem counterintuitive. Many people might be surprised to learn, for example, that the adult literacy rate in Mongolia is identical to France's, or that Cuba's infant mortality rate is the same as that in the United States. It seems likely that the information people consider, or the inferences they make, suggest large inequalities where none exist. While Cuba's per capita income is thirteen times lower than the United States, this has no impact on Cuba's infant mortality, life expectancy, or literacy rates compared with the U.S. This is not to suggest that wealth is a poor predictor of quality-of-life statistics, rather that item level differences can occur for many reasons other than wealth.

Sources of error can often be traced to the estimation strategies people select. For example, a strategy for judging event frequency is to recall each occurrence as it comes to mind and count each of the recalled instances. Consequently, the enumeration strategy is mediated by memory and can result in underestimation because forgotten events are omitted (Blair & Burton, 1987; Brown, 1995; Brown & Sinclair, 1999).

Systematic errors, or biases, can also be associated with specific types of numerical domain (e.g., population, infant mortality, prices, dates, land area, and distances). While populations tend to be underestimated, infant mortality rates are most often overestimated. These differences are frequently due to the way that respondents select information and how they infer a range of possible responses. The picture is complicated further because people frequently consider more than one strategy and different types of information when answering a single numerical question. As a result sources of error can differ within the same subject and within the same numerical domain (Brown, 2002).

Despite a tradition of focussing on different estimation strategies (e.g., Gigerenzer, Todd, & the ABC Research Group, 1999; Kahneman, Slovic, & Tversky, 1982; Reder, 1982, 1987; Tversky & Kahneman, 1974), a newer approach has been to examine how numerical knowledge is represented and the effects of information selection on accuracy (e.g., Brown, 1995; Brown & Siegler, 1993; Friedman, Brown, & McGaffey, 2002). The relationship between information and strategy is important, and may impact accuracy for two reasons. First, people's misconceptions can introduce errors that may lead to impoverished notions about either the numerical domain in general or the specific item being estimated. For example, neither geographical nor population size have any impact on quality-of-life statistics across all countries in the world, although extremely large populations may exacerbate high levels of poverty due to the absolute magnitude of the social and economic problems involved. As a result, India and China may be considered as developing countries, but the difficulties these countries have bringing, for example, adequate health care to all of their citizens may often go unappreciated. Secondly, the strategies that people select are often bounded by the content and availability of information that can be retrieved from memory (Brown, 2002), so that an absence of domain or item specific knowledge can result in guessing strategies rather than estimates grounded on plausible reasoning (Lee & Brown, 2004).

Metrics and Mapping

Metrics and mapping (Brown & Siegler, 1993) provides a framework for understanding how different types of information impacts on estimation accuracy. This conceptualization illustrates how metric and ordinal information tends to influence metric and ordinal accuracy independently. *Numerical-retrieval* describes one class of metric strategies that rely on metric information, such as knowledge about a domain's unit of measurement and its statistical properties (e.g., the range and central tendency). The ability to recall metric information or learning new information can have a significant impact on metric accuracy (as measured by the absolute deviation of estimates from their actual values). Furthermore, metric knowledge permits respondents to focus their attention on generating absolute numerical values.

Ordinal-conversion, on the other hand, refers to a set of processes that are evoked by information about the rank ordering of items. Factors that influence ordinal accuracy (measured by the rank order correlation between estimated and actual values) have been shown to act independently of those that act on metric accuracy. Because these different classes of information have dissociated effects on different types of accuracy, the metrics and mapping approach stresses the importance of understanding information types and the cognitive strategies that are associated with them. Furthermore, these relationships are informative because response types, and patterns of accuracy, reflect how individuals represent, organize, and use different types of real-world information to generate plausible estimates (e.g. Brown, 2002; Friedman, Kerkman, & Brown, 2000).

The following research project focuses on ordinal-conversion, and extends our understanding, in part, by investigating how one type of information called *proxy variables* could be the primary source of information used to generate ordinally ranked numerical estimates.[1] Investigating these processes is important because people often know little about the actual metrics they are being asked to estimate, but are still able to resort to strategies that permit them to rank order items. For instance, it is generally known that Mars is closer to the Sun than Saturn. However, few people could confidently report that their mean distances are 140 and 880 million miles respectively, although their numerical estimates would very likely be consistent with their ordinal understanding.

Because ordinal relationships may be more readily recalled than metric characteristics, ordinal-conversion is assumed to be the default method for answering most types of numerical question (at least for non-experts). When there are only a few items in a set, such as planets in the solar system, these relationships may be easily learnt and recalled, even if the actual metric values are not. Still, in domains where there are many potential target items (such as countries), rank orders may be too numerous to be recalled accurately or may be unknown. This is not to imply that ordinal judgments are no

[1] Proxy variable refers to a domain of ordinally predictive information whereas *proxy cue* designates specific instances of the information type. For example, if land area were a proxy variable then India's land area would be a proxy cue. The term *proxy/ordinal-conversion hypothesis* refers to the testable notion that proxy cues are a primary source of ordinal information operated on by ordinal-conversion processes.

longer possible. Proxy variables can be used to generate inferences about the approximate relationship between items on a scale. For example, it could be correctly inferred that because Egypt is considerably poorer than the developed nations that the Egyptian literacy rate is likely to be low rather than high. A proxy/ordinal-conversion hypothesis argues that in the absence of numerical facts people will ask themselves what else they know about a target item, and draw inferences based on the relationship between target items and subjective values on a proxy variable (Broniarczyk, & Alba, 1994; Dawes, 1998; Simon, 1956; Simon, et. al., 1986).

There are strong theoretical grounds for assuming a pivotal role for proxy variables in numerical estimation, and that plausible inferences can be made using proxy cues. One line of support is derived from paired comparison tasks, where two items have to be judged. For example, estimating whether the population of Los Angeles is greater or less than Montreal's. Gigerenzer, Todd, & the ABC Research Group (1999) have argued that paired comparisons can be accurately generated using one-reason decisions based on single predictive cues, such as the *recognition* and *take-the-best* (TTB) heuristics. In the former, recognition is used as a cue, so that recognized cities, for example, are judged as having larger populations than unrecognized ones. This kind of plausible reasoning is principled, because large cities are more likely to be encountered (e.g., in the news) than smaller cities. This is to say that the heuristic relies on the ecological validity that people are more exposed to what happens in larger cities than smaller ones. TTB on the other hand assumes that the single most predictive cue capable of resolving a binomial decision is used as the sole source of information. The hypothesized process is multi-staged. First cues are rank ordered according to their predictive power, from highest to lowest. The most predictive cue is chosen first and a comparison is made between the two target items. If the most predictive cue is not capable of resolving the decision, then the next most predictive cue is chosen, and so on. This heuristic of cue selection has many potential advantages. For example, by capitalizing on the strongest available relationship between what is known and the to-be-estimated value, but at the same time rejecting cues that are incapable of determining a solution. The main points of interest are that both heuristics rely on ecologically valid information and that binomial decisions are similar to rank ordering.

QuickEst

Focusing on how people make ordinal judgments rather than paired comparisons, Hertwig, Hoffrage, and Martignon (1999) have proposed a heuristic called QuickEst. This heuristic is derived from a model simulating numerical estimation processes, using as little information as possible. This heuristic is believed to be fast and frugal because it models processes that try to access the minimum number of ecologically valid cues capable of determining a solution. A second feature is that numerical estimates are thought to reflect a series of modal numerical values (e.g., Baird, Lewis, & Romer, 1970; Huttenlocher & Hedges, 1994). This means that for a given domain there exists a set of numerical categories, called *spontaneous numbers*, which capture target items that have shared characteristics in a group (i.e., membership and typicality).

QuickEst has also been characterized as a heuristic for estimating target items that demonstrate J-shaped frequency distributions. These functions describe frequency plots where there are more items at one end of a scale so that the frequency density decreases monotonically toward the other end of the scale. The model capitalizes on this type of structure in the environment because the information search initially focuses on cues that are able to distinguish the most frequent items on the J-function from other items. An example provided by Hertwig, et al. (1999) is that few small cities in Germany have a professional soccer team. Hence the presence or absence of a professional soccer team is a good cue for estimating the populations of German cities. Furthermore, the cue is ecologically sound because small cities are unlikely to attract enough support to finance a professional soccer club.

Having distinguished a target item on the basis of category membership, the next step is a search for another cue that continues to discriminate the target object from some of those that still remain. This process continues until the search for information stops. This occurs when the target item is found to *not* to possess the feature associated with the cue. For example, when estimating city populations, if a city possesses the first few features in hierarchical order (such as a soccer team, and international conference center) but does not possess, for instance, a large concert hall then the search will be terminated on the concert hall cue. Once information search has been terminated a numerical estimate can be generated. People are assumed to decide on a specific numerical value by providing a mean value on the target domain (or criterion variable) that reflects the items category assignment (e.g., a mean value associated with cities with soccer teams and conference centers, but without concert halls).

The underlying characteristic of the heuristic relies on eliminating the common criterion categories, and using the absence of features to remove less common categories, termed *elimination by aspects*. As a result one category remains that distinguishes the target item (and similar items) from all other categories. The most important feature is the absence of any cue combinations; although multiple cues may be used (depending on the actual value of the to-be-estimated item on the criterion variable), no cues are combined and thus no cue weighting or Bayesian reasoning is applied. Despite claims that QuickEst is fast and frugal this thesis will show that more parsimonious solutions can be achieved, and that an empirical understanding of estimation strategies is critical to understanding how information is selected and used.

Still, QuickEst and the proxy/ordinal-conversion positions share one underlying principle. Both positions rely on the assumption that ecologically valid information is relevant and can be recognized by respondents because criterion and predictor variables co-occur in the environment with enough frequency that their causal relationships can be learnt. However, there are also fundamental points of departure. QuickEst does not distinguish between different types of information (only its ability to assign target items to categories) while the proxy/ordinal-conversion position posits metric and ordinal information as fundamentally different types of information. Furthermore, the proxy/ordinal-conversion approach assumes (as will be explained further on) that estimation strategies can be modified when proxy cues are non-existent or fail to predict values on the criterion variable. Lastly, because Quickest makes no distinction between metric and mapping information their independent effects on metric and ordinal accuracy is overlooked. Surprisingly, supporting evidence for QuickEst has focused on metric accuracy (Hertwig, et al., 1999) and not ordinal accuracy where QuickEst should have the most impact because accurate assignment to categories on the target domain should result in a fair degree of ordinal accuracy (depending on how many spontaneous categories are associated with the target domain). Hogarth and Karelaia (in press) have tested mathematical simulations that generally uphold the claim that one-reason decisions could be as effective as multiple cue solutions, at least when there is a high ecologically valid correlation between the cue and target variables (i.e., high ecological validity).[2]

This is not to say that the proxy/ordinal-conversion hypothesis assumes that people always identify and using the single most predictive cue, or that proxy cues are the only information accessed under all conditions. The final experiment presented here will show that the proxy/ordinal-conversion position is flexible enough to include the use of additional information (under certain circumstances); moreover that cognitive adaptability is a critical attribute of real-world numerical estimation that can be characterized as flexibility between strategy and both the quantity and type of information that can be applied. As a result, the position does not claim that single cues will invariably provide the only solution when people make numerical inferences, compared with the multi-cue position taken by QuickEst.

On this view, this program of research addresses some of the questions raised regarding the validity of single reason decision-making in real-world tasks. Newell and Shanks (2003) have questioned the psychological validity of "take the best" and point to a lack of empirical evidence for one-reason decision making as a general phenomenon. These researchers point out that the cost of information, knowledge about cue validities, and size of the problem space are critical factors that determine the likelihood that

[2] Throughout this thesis ecological correlation refers to Brunswik's (1952) conceptualization, and is not related to the concept of external validity put forward by Campbell & Stanley, (1966) which refers to generalizable research methods.

respondents will respond on the basis of single cues (see also Broder, 2000; Broder & Schiffer, 2003; Dhami & Ayton, 2001; Newell, Weston & Shanks, 2003). Despite the apparent juxtaposition between single versus multiple cue use, the proxy/ordinal-conversion approach does not attempt to distinguish one-reason responses from other positions. Rather, it demonstrates a general applicability of the ordinal-conversion strategy and the role of proxy variables as an important (if not always the only) source of task relevant information. Nevertheless, this viewpoint follows the current consensus that plausible reasoning may not provide optimal solutions in estimation under uncertainty, but may be adequate for generating credible solutions when time and availability of task relevant information are constrained, e.g., *satisficing* (Simon, 1956, 1969) and *bounded rationality* (Simon, 1955, 1979). Satisficing refers to the idea that people use enough information to generate plausible solutions rather than attempting to find optimal solutions. Bounded rationality, is similar conceptually, and refers to human cognition as constrained by time and available information (although these description are not definitive, they provide enough conceptual clarity to frame the current project).

Still, it needs to be empirically established if people use proxy variables to generate numerical estimates. In addition, to provide a principled investigation of the proxy/ordinal-conversion hypothesis must also describe how ordinal-conversion can be deconstructed into sub-processes to explain how information is searched and how numerical responses can be generated from proxy cues.

The next section provides a detailed exposition of ordinal-conversion and the sub-processes that characterize how people search for information, how they use information to infer numerical quantifiers and plausible ranges, and how they finally generate numerical values from their ordinal inferences. To describe how proxy cues might be usefully employed as ordinal information it is necessary to understand each of these sub-processes. It is also relevant to consider that the same proxy cues can be called upon repeatedly (i.e., by more than one sub-process), and that additional information may also be retrieved when proxy cues lack discriminatory power. In the final section of this chapter, the theoretical characteristics of proxy variables will be described, showing how they differ from other types of information. Finally, the chapter will outline a series of testable hypotheses that support the claim that people may rely on proxy information when making certain types of numerical estimates, and thus provide the rationale for Experiments 1 and 2.

Ordinal-conversion

Information Search

People may consider a broad range of information when generating numerical estimates. While some information will consist of verifiable facts other details may reflect non-verifiable hearsay, such as anecdotal evidence.[3] Still, it is assumed that the search for task relevant information is opportunistic because respondents have little idea, a-priori, whether the details they will retrieve have any substantive merit. It is also assumed that information is evaluated immediately when it comes to mind because serial evaluation has the advantage of limiting working memory load, does not require mechanisms capable of coping with indeterminate amounts of information and, most

[3] The notion that people rely on different sources of information is not meant to imply that verifiable information is necessarily better than anecdotal evidence.

importantly from a theoretical standpoint, avoids the need to combine cues, i.e., problems associated with Bayesian reasoning such as failure to take into account base rates (Kahneman & Tversky, 1973). If the information is judged as credible it may lead to an inference about the target item and it's ordinal position on the criterion variable being estimated, otherwise information and inference are rejected and the search process continues iteratively in a *retrieval-inference cycle*. Thus, the goal of searching memory is to bootstrap a plausible inference from a single piece of information.

Unlike contemporary models of decision-making, for example the "fast and frugal" approach (Gigerenzer & Goldstein, 1996) and QuickEst in particular, the retrieval-inference cycle requires no formal stopping rule. Instead, the successful generation of a plausible inference or motivation and time constraints determine when the retrieval-inference cycle is terminated (Payne, Bettman, & Johnson, 1993). For example, if no appropriate inferences are elicited the respondent may resort to guessing by engaging processes that require no inference about the target item (Lee & Brown, 2004), or respondents may simply give up and not respond. However, when successful, the metrics and mapping framework assumes that if the inference is ordinal (i.e., refers to the items relative position across the set and with other specific targets), then ordinal-conversion procedures will be evoked. These procedures describe a set of independent serial processes called *setting the metric / determining the range, partitioning the range, assigning target items to partitions,* and finally *numerical selection*. In the remainder of this section, each of these procedures will examined in turn, with a focus on how ordinal-conversion can capitalize on real-world knowledge such as proxy cues.

Setting the Metric / Determining the Range

It is certain that all numerical estimates must embrace one or more quantitative decisions. There is no basis for making a numerical response without first deciding on an appropriate unit of measurement in which to respond or establishing the range of numbers likely to be involved. Setting the metric describes how people choose a metric, and the likely units of measurement (e.g., miles, percentages, volts, tens, thousands, millions). Sometimes the metric component is given in the question, allowing this process to be bypassed, e.g., "What is the population *(in millions)* of Peru?" In more complex situations the metric is not always obvious and may have to be decided by the respondent without the aid of external prompts.

Once a metric has been chosen it becomes possible to determine a plausible range of numbers by postulating upper and lower boundaries capable of capturing responses from across a possible set of items. However, ordinal-conversion assumes that there is no optimal method for determining the range that can be applied to all numerical domains. In some cases people may rely on metric knowledge for a few known items and attempt to infer boundaries based on these values. Population estimates, for instance, often seem to rely on knowledge about the population of the respondents' own country and knowledge that China has the largest population with approximately one billion inhabitants. These facts can provide important information suggesting that populations are likely to be measured in tens or hundreds of millions. On the other hand, people rarely know any numerical values for infant mortality or literacy rates. Unless the upper and lower bounds can be recalled directly, the range may have to be inferred using common sense and intuition. For example, infant mortality rates cannot be zero because it is inevitable that some children will die. Similarly, it can also be inferred that mortality rates will unlikely be so high as to expunge an entire nation of people.

In addition to providing the basis for making numerical responses, setting the metric and determining the range also serves as a framework for maintaining internal consistency across responses because each response can be generated using the same metric and range. This is not to suggest that people are unable or unwilling to adjust their metric or range, although updating typically occurs when new numerical information is encountered that cannot be accommodated by people's existing metric and range beliefs (Brown, 2002).

Setting the metric and determining the range are hypothetical procedures that have not yet been tested directly. Nonetheless, these are assumptions are grounded on a substantial body of evidence that people employ metric and ordinal information independently (Brown, 2002; Brown & Siegler, 1996), and a principled argument that changes in performance can be attributed to updating metric knowledge. A three-phase testing procedure called seeding has been used extensively to show how different types of information impacts performance (Brown, 2002; Brown & Siegler, 1996, Friedman & Brown, 1999, 2000a; LaVoie, Bourne, & Healy, 2002; Lawson, & Bhagat, 2002). In the standard seeding paradigm, participants are first asked to estimate the values for a set of target items. These responses provide the initial estimates that can later be compared using a within subjects design. Participants are then shown the true values for a sub-set of items, called *seeds*. Seed information provides new metric information that allows respondents to update their metric beliefs and readjust their metrics and range if necessary. Finally, the participant is asked to re-estimate the complete set of target items for a second time. The second estimates for the non-seeded items can then be compared with each respondents' initial estimates, and the impact of accurate new information on accuracy can be measured. Seeding has provided extensive evidence that metric accuracy for the non-seed items invariably increases at retest (cf. Murray & Brown, manuscript under review) due to respondents re-evaluating their metric understanding in light of information that suggests their original beliefs about the magnitude and range of values was incorrect.

Partitioning the Range

Setting the metric and determining the range are necessary components in a model that explains how people use ordinal information to make numerical estimates. Without establishing a common metric and the range of possible values, target items cannot be compared with other items. Determining the range leads to further processes that narrow the range of plausible responses, by partitioning the range into categories. Each partition captures a potential subset of values that can be mapped by descriptive quantifiers, such as "high", "medium", or "low" categories. This does not imply that respondents always rely on three-category representations, or that respondents use the same number of categories, so the number of partitions reflects individual representations about the numerical domain.

Still, some numerical domains may have natural categories. For instance, sports teams are often categorised into divisions or leagues that classify teams with relatively similar abilities. More concretely, Friedman and Brown (1999, 2000a, 2000b) have shown that North Americans' estimates for the latitudes of cities in North America cluster into four distinct regions, although their target stimuli had been selected uniformly throughout Canada, the U.S. and Mexico. These data indicate a four-partition representation based on psychological boundaries, over and above the three-partition solution suggested by geopolitical boundaries. Similarly, price estimates for automobiles have been associated with two-cluster representations as respondents tend to partition estimates into high and low prestige vehicles (Murray & Brown, manuscript under review).

Proxy variables may be a particularly important source of ordinal information when partitioning ranges. This is because the same predictive information that correctly suggests, for example, that the infant mortality rate in Sweden is lower than in Somalia, can also be used to infer how nations might be categorized and the ordinal relationship between partitions (e.g., Sweden, Japan, and the U.S. vs. Somalia, Afghanistan, and Laos).

However, the use of proxy variables may also be a source of estimation error. Proxy variables are subjective and may not result in veridical partitions being selected, and respondents will likely to be unaware of how accurately they have partitioned the range based on proxy variables. Still, they should still be able to make principled ordinal numerical judgments because the relative ordering of partitions should not be affected by metric inaccuracy.

Numerical Selection

The final process in ordinal-conversion is the selection of a numerical value that reflects the target items ordinal ranking relative to other items. This can be achieved by assigning the target item to a partition, and then deciding on a plausible quantity that reflects the item's relative rank. Like QuickEst, items can be assigned to partitions on the basis of category membership. Again, proxy cues may provide relevant information for the same reasons that proxy variables enable people to partition the range, although this time proxy cues are used at the item level. For example, if wealth were used as a proxy variable to estimate infant mortality rates, then France, Israel, and the U.S. might be categorized as wealthy, and members of a low mortality rate partition.

Respondents are now able to decide on a numerical response. Unlike QuickEst, it is not assumed that items assigned to the same partition will also be assigned the same numerical value (i.e., a number equivalent to the subjective mean of items in the partition). This means that participants might assign France, Israel and the U.S. to the low infant mortality partition, but ordinal-conversion assumes it is unlikely that participants will respond by giving the same estimated mortality rate for all three countries. Although it is difficult to maintain ordinal consistency at the level of individual items, individual responses can be made in two ways using proxy cues as the primary source of ordinal information.

The first method is to select a numerical value that reflects the item's ordinal position on the proxy variable in relation to the target item's immediate neighbours on the same proxy. The fact that the per capita wealth in the U.S. is higher than Israel's correctly predicts that the U.S. mortality rate should be lower than Israel's. Using proxy cues, the respondent can select a numerical value from the partition that is consistent with the perceived ordinal rank on the proxy variable. However, using proxy cues in this way might also be a source of inaccuracy because the same assumptions do not hold between, for example, France and the U.S., where American children are almost twice as likely to die despite a significant U.S. advantage in overall wealth. This type of anomalous example reinforces two points. Firstly, that ordinal-conversion is an example in satisficing because proxy variables may be imperfect but generalizable correlates that describe trends rather than optimal solutions.[4] Secondly, proxy cues may not be the only source of information available; respondents may be able to identify instances when proxy cues are not predictive at the item level and may also recognize other, more important, factors that determine a country's infant mortality rate.

Therefore a second method of selecting numerical values is posited. This method adjusts the final numerical estimate, constrained by the range of values suggested by the assigned partition. If the proxy cue cannot discriminate between items from items assigned to the same partition, then the final estimate would be selected from the range possible range responses associated with the target items partition. In order to achieve this, other item specific information would need to be recalled capable of generating plausible reasons for distinguishing the target item from other partition members. Returning to the France/U.S. anomaly, the inference that France's mortality rate is higher than in the U.S. (based on a disparity in wealth) might be rejected if other information can also be retrieved. For instance, if the fact that forty million Americans live without healthcare coverage can be retrieved and that France has a national healthcare system, implies the opposite, that Americans' are likely to be in poorer health than French

[4] This is not to claim that anomalies automatically result in ordinal inaccuracy. Estimation accuracy is likely to depend not only on the strength of the perceived relationship between proxy and target variables, but also the face validity of the proxy cue.

citizens. In other words, if other pertinent facts can be recalled then some adjustment can be rendered that accounts for both the proxy information (i.e., that both France and the U.S. both have low mortality rates) and access to health services (i.e., that the U.S. has a higher mortality rate than France).

Still, the probability of using this second method for selecting numerical values is likely to depend on individuals identifying when the proxy cue is a poor predictor at the item level and whether the same individual can retrieve additional item specific information. If no additional information can be recalled the final numerical judgment will be based on the subjective value of the proxy cue alone.

It follows that for a given set of target items, there are likely to be significant individual differences in the number of cues used. If some respondents use more than one cue, it does not appear appropriate to characterize proxy variables solely on the basis of their rank order correlation with the criterion variable. This is because the subjective value in using proxy cues depends on individual differences in both item and domain specific knowledge (i.e., not all respondents will use the same cues, irrespective of the actual validity of the cues). The next section describes three criteria that help differentiate proxy variables from other information, and outlines how asking respondents to estimate quality-of-life measures might be used to test the proxy/ordinal-conversion hypothesis.

Proxy Cues and Variables

Many of the preceding examples, and all of the following experiments, refer to infant mortality rates as the to-be-estimated domain. Quality-of-life statistics are useful for describing the nature of proxy variables because they typify conditions where the ordinal-conversion process is likely to be most effective. More importantly, quality-of-life statistics, such as infant mortality rates, life expectancy, and literacy rates are well predicted by a single proxy cue: wealth.

It is assumed that most people understand the concepts represented by quality-of-life statistics, although the actual metrics used to describe may be poorly known. For example, people are likely to know that infant mortality rates reflect children's well-being and the probability of their survival without knowing that infant mortality is measured as the number of children who die, per one-thousand live births, before age one. Because few people are likely to know how infant mortality rates are measured it follows that they will also be unaware of the statistical properties of the domain, such as the range and central tendency. Nonetheless, proxy variables may still provide the type of ordinal information that can predict one country's infant mortality rate as lower than another, on the basis that rich countries are very unlikely to have higher infant mortality rates than poor countries, even if the actual values are unknown.

Still, proxy variables need to be distinguished from variables that are simply predictive. It is assumed throughout this thesis that proxy variables differ from other variables whose only relevant feature is that they correlate with the criterion variable when the proxy variable (predictor) is more familiar and better known than the target domain (criterion), and there is a perceived direction in the relationship between variables.

Proxy variables need not, however, be veridical (i.e., the assumed characteristics may be subjective, rather than objective). Provided a variable is perceived as predictive in a known direction and is better known than the criterion then the cue may be considered (rightly or wrongly) as proxy information. This caveat accepts that people engage in ordinal-conversion processes even when the information they retrieve has little or no actual predictive power. In these cases, it should be possible to measure when respondents fail to identify appropriate cues which may lead to errors that can be identified and measured using process tracing techniques (see Svenson, 1996), although this suggestion is beyond the scope of the current project.

I have suggested in this chapter that wealth might be a good proxy variable for estimating quality-of-life statistics in other countries. This is because wealth is assumed to be conceptually better understood than infant mortality, literacy and life expectancy figures. Across one hundred and forty seven countries, wealth (measured by per capita gross domestic product, or *pcGDP*) is a strong predictor of infant mortality ($r_s = -0.87$) and comprehensible enough for the direction of the relationship to be discernable. Wealth might also be an effective proxy variable for estimating human life expectancies ($r_s = 0.84$), literacy rates ($r_s = 0.72$), or levels of malnutrition ($r_s = -0.78$). Pilot studies have shown that when asked to estimate pcGDP for 92 countries from around the world respondents are able to do so with a relatively high degree of ordinal accuracy ($r_s = .63$). These data suggest that these respondents appear reasonably familiar with the concept of national incomes and the rank ordering of countries. In contrast, the population of a country, for instance, would not be a reliable cue to estimate quality-of-life statistics (max. $r_s < .05$).

In the next chapter I test the assumptions that proxy cues are a modal source of ordinal information, and that people are able to identify proxy cues, not only as good predictors, but also causally related in a perceived direction. Moreover, Experiment 2 investigates the types of information people use to estimate infant mortality rates and the relationship between actual wealth and estimated infant mortality rates. If people are capable of utilizing the predictive validity of proxy variables, then wealth should be considered a primary source of task relevant information when estimating infant mortality rates. Nevertheless, other cues should also be used, reflecting differences individual in real-world knowledge and differing beliefs about cue validities.

It is also predicted that a general characteristic of infant mortality estimates is that they will tend to be overestimated. This is because mortality rates span less than one-fifth of the scale that is used to measure them. As a result, respondents will fail to partition the range accurately due to insufficient numerical knowledge. However, if respondents rely on ordinal-conversion and proxy cues, such as wealth, then ordinal accuracy should be high due to the strong ecological validity between wealth and infant mortality rates.

Chapter II

This chapter presents two exploratory investigations that examine the assumptions underpinning the proxy/ordinal-conversion hypothesis. To begin, it is important to establish whether people are able to identify proxy variables. If people are unable to identify highly predictive cues there is little basis for assuming that proxy cues are a primary source of ordinal information in numerical estimation. Experiment 1 is an exploratory investigation into people's beliefs about differences in quality-of-life statistics in different countries, and whether these beliefs qualify as proxy variables as defined in Chapter I. The claim that overestimation should occur when actual values do not span the scale used to measure them is also examined.

Second, it is necessary to investigate whether proxy cues serve as a source of ordinal information when generating numerical responses and the relationship between actual pcGDP and estimated infant mortality rates (Experiment 2). Experiment 2 also focuses on the frequency that respondents report different types of cues when estimating infant mortality rates. Because metric and mapping information is assumed to be independent, this study also examines differences between respondents' metric and ordinal accuracy.

Experiment 1

A core assumption of this research program is that people can frequently identify proxy variables, even when generating estimates for numerical domains that they know very little about. Theories that emphasize the utilization of limited amounts of predictive information (e.g., the proxy/ordinal-conversion and "fast and frugal" positions) make two broad claims regarding the way that a) information is structured in the environment and b) how information is accessed in memory. The first claim is that the environment reflects causal relationships. For example, few people would doubt that in the natural world poverty may result in famine in times of drought. This is to say that the relationship between famine, disease, and their impact on infant mortality is structured in our understanding of the natural world. The second claim is that causal phenomenon and their effects are encoded, so that the retrieval of task relevant information can lead to inferences about the relationship between proxy variables and associated target domains.

Despite the fact that much of the psychology of human memory is grounded on principles of association, it is necessary to clarify exactly why co-occurrence provides ecologically valid information structures, and how the structure of information could aid the retrieval of cues. One theory is that causes and their effects, for a given set of occurrences, are likely to be better recalled when causes generalize across multiple instances as opposed to causes that only result in effects for a subset of instances (Allan, 1993; Cheng & Novick, 1990; Shanks, 1995). This is presumably because generalizable causes are encountered more often and are therefore more readily available. The probabilistic contrast model posits that causal induction is normative, and that associations between causes and their effects can be learnt from associations in the environment that are necessarily contiguous, so that the cause necessarily precedes the effect. The model also proposes that biases (i.e., respondents selecting non-predictive over predictive information) are individual differences with subjectively coherent reasons and not departures from normative processes. In terms of cue selection, the model predicts that the use of non-predictive cues results from beliefs in the predictive power of the cue rather than differences in cue selection strategies. This is contrary to the Rescorla-

Wagner model of associative learning (Rescorla & Wagner, 1972) that suggests associative learning is more likely to occur when unusual or surprising associations are encountered. The proxy/ordinal-conversion hypothesis assumes that surprise is inversely related to the probability that a proxy cue will be recalled, so that pairs associated with the highest covariance (i.e., the most common and least surprising) will be judged as better predictors and more causally related. The relationship between the probability that a proxy cue will be accessed can be modelled using Equation 1. This equations states that the probability of a proxy cue being selected as the primary source of ordinal information is the probability of cause and effect being both present minus the probability of the effect occurring in the absence of the same cause (i.e., the strongest associations are least surprising because contrary examples are less frequent than for surprising associations).

$$\text{Prob. proxy accessed} = P[\text{Effect}|C] - P[\text{Effect}|{\sim}C] \tag{1}$$

* where P = prob. the effect + cause is present (|C), or absent (~C)

This position predicts that wealth is more likely to be recalled and also judged as a better predictor of infant health than, for instance, conflict; in the real-world high infant mortality rates always co-vary with high levels of poverty, while conflict does not (wealthy countries conduct wars without unduly influencing their own infant health though not, perhaps, their poorer enemies). While ecological validity determines how predictive a cue is, the probability that a predictive cue will be used is a mnemonic phenomenon, but one that is mediated by frequency rather than surprise. The claim that ecologically valid information is structured in the environment draws upon a broader conceptualization of cognition as primarily goal directed (Gigerenzer, et. al., 1999). That is people first receive then process information and learn to react intelligently when the same (or similar) environmental factors are encountered (Brunswik, 1952).

The second claim (that the retrieval of associated causes can lead to plausible inferences) is related to the first. People's memories hold a range of information that could be used to generate any number of inferences in pursuit of a multitude of different answers. On one hand, some of the information people recall will not result in a usable inference and the retrieval-inference cycle will continue to search for task relevant information until a plausible inference can be made. On the other hand, it is psychologically implausible that people interrogate all of the information that they could potentially access. How is it that people can have credible grounds for having faith in a relatively small amount of information, or infer that one type of information from this subset is more relevant than another? As an illustration, consider the following description (each item of information having been numbered):

> "Bangladesh came into existence in 1971(1) when Bengali East Pakistan seceded from its union with West Pakistan(2). About a third of this extremely poor(3)country floods annually(4) during the monsoon rainy season(5), hampering economic development (6)." (CIA World Factbook, 2004).

This is a credible representation of the type and quantity of information a reasonably well-informed person might retrieve about Bangladesh. The argument that ecological valid cues are structured in the environment suggests that items 3, and 6 are pertinent for estimating Bangladesh's infant mortality rate. Nonetheless, there are also two groups of information that do not seem directly relevant (especially if more predictive information can also be recalled). Items 1 and 2 are useful as historical references, while 4 and 5 refer to geographical characteristics. One of the problems for theories of real-world estimation is to explain how people search out likely sources of information while ignoring others.

Ordinal-conversion assumes that questions about a country's infant mortality rate contains two probes; an implicit probe that relates to, and restricts, search to information pertaining to infant mortality rates in general, and a second explicit probe that interrogates memory for facts about a named target country. Combined, these probes are likely to result in the retrieval of proxy information, but are unlikely to bring irrelevant items to mind, such as Bangladesh's independence. A combination of two probes circumvents the need to marshal all of the known facts, or the need to disregard information through a process of serial elimination (cf. Hertwig, et. al., 1999).

Successful retrieval of the most predictive cues is far from certain, as is the ability to generate plausible inferences. Rather than being a brittle description of plausible reasoning, the ordinal-conversion process may still lead to principled solutions even when proxy variables cannot be identified, or proxy cues are absent. Inferences can still be made, albeit grounded on less relevant information. For example, Bangladesh's infant mortality rate could still be loosely inferred based on a particularly harsh climate, or by association to Pakistan (despite the fact that these cues are neither generally predictive, nor reliable). Of course, if nothing at all is known about the target then guessing is likely to be the only viable estimation strategy. Still, the dual-probe model provides a flexible architecture for retrieving information without recourse to combinatorial cues or elimination by aspects.

The following experiment provides an empirical basis for testing the preceding arguments and to establish whether people are able to associate a veridical proxy variable with a target domain. In Experiment 1 respondents estimated the infant mortality rate for a country. They were then asked to describe the most important factor that influences quality-of-life such as infant mortality, literacy and life expectancy for countries in general. If respondents can readily access task relevant proxy cues, then their modal response to the question about the most important factor influencing mortality should be references to wealth and wealth related factors, e.g. finance, commerce, technological advancement, and standard of living. A secondary prediction is that estimates should be overestimated due to errors in determining an accurate range of responses. Because the highest infant mortality rate at the time was 147 deaths per 1000 live births, participant's subjective upper boundary was likely to be higher than the true upper bound, with a corresponding tendency to partition the range in such a way that all items are likely to be overestimated.

Method

Participants

Two hundred and eighty seven undergraduates from the University of Alberta participated (157 females, 119 males, and 10 unspecified). Eighty-five percent were Canadian born, with a median age of 19 years. Participants were tested together in classrooms prior to the start of class and received course credit for their involvement. This was a between subjects questionnaire where participants were randomly assigned to answer a single infant mortality question for a target country selected from a set of twenty possible targets.

Materials and Procedure

Twenty countries (Appendix I) were chosen from a database of 146 nations. The rank order correlation between pcGDP and infant mortality for these stimuli was representative of the relationship between pcGDP and the database as a whole ($r_s = -.86$). Pilot research indicated that these countries were often recognized and likely to elicit inferences rather than guessed responses (mean probability of recognition > 97%). The selected countries spanned the true range of infant mortality rates (3/1000 for Japan to 147/1000 in Afghanistan). Paper booklets were constructed, consisting of a cover page asking for the participant's details including their country of birth. The participants were informed that the research was directed at how people estimate certain types of social statistics. They were instructed to complete the booklet moving from the cover page to the back page in order.

On the second page, participants were told that they would be estimating infant mortality and were given the name of their target country. It was explained that infant mortality is measured by the number of children who die, prior to their first birthday, for every one thousand live births. It was pointed out that their responses should therefore be between zero and one thousand, inclusive. They were then asked to estimate the infant mortality rate for the target country.

Finally, on page three participants were asked to describe the single most important factor that affects people's health and education in other countries around the world (not just the target country they had estimated). They were told that "factor" meant the one element that determines life expectancy and literacy rate as well as infant mortality in different countries. The participants were instructed to provide brief answers to this question, i.e. one or two words.

Results and Discussion

One participant's mortality estimate was outside the specified boundaries and was excluded in the following analysis. Two analyses were then conducted. The first analysis examined categories of answers participants gave to "most important factor" question and their frequencies. A second analysis examined metric accuracy and biases for the infant mortality estimates. Because participants responded to only one country, it is not possible to analyze rank order accuracy. However, it is possible to examine mean metric accuracy for each item (i.e., how far the numerical estimates for each item deviated from their actual values).

Accuracy

Two measures of accuracy were calculated. Signed order of magnitude error (*SOME*) is used to determine estimation bias. A SOME score was derived for each participant by calculating the $\log^{10}$ of each estimate after dividing the estimate by the country's actual infant mortality rate. Positive figures indicate overestimation and negative figures signify underestimation. Order of magnitude error (*OME*) measures absolute accuracy, irrespective of bias, and was determined by calculating the absolute value of the obtained SOME score for each participant. These measures are convenient for interpreting large error rates, and for comparing error rates across items. For instance, a mean SOME of 1.0 indicates that responses were overestimated by an average of one order of magnitude from the actual values (e.g., estimating Russia's mortality rate as 200/1000 when the actual value is 20/1000). When averaging over SOME values, the effect of positive figures will offset by negative values. Mean OME, on the other hand, disregards differences between biases and describes the mean absolute deviation from the actual values.

Overall, respondents overestimated infant mortality rates, mean SOME = .68 (*SEM* = .04), with 89% of participants providing estimates larger than the actual values of their target countries. Over estimation was predicted because the potential response range is wider than the true range of infant mortality (between one and approximately one hundred and fifty deaths per thousand live births). The proxy/ordinal-conversion position assumes that the primary strategy for estimating infant mortality rates involves accessing ordinally relevant information. Because inferences generated from the retrieval of subjective wealth information are ordinal inferences, a lack of understanding for the metric properties of the domain may lead to inaccuracies in partitioning the range. Absolute metric accuracy was also poor, mean OME = .76 (*SEM* = .03). This supports the assumption that subjective proxy information, such as wealth, does not provide accurate metric information. A pilot study where respondents estimated pcGDP for the same set of countries, people were less biased when estimating pcGDP (mean SOME = 0.01, *SEM* = .03) and more accurate (mean OME = 0.66, *SEM* = .02). Although it is problematic comparing the results from different studies, the large numbers of participants tested in both experiments suggests that the difference in estimated parameters is stable and supports the hypothesis that proxy variables are better known than their corresponding criterion variables.

Causal Explanations

This was a four-phase analysis loosely following the principles of grounded theory (Glaser & Strauss, 1967). Because there was no a priori framework for coding these responses the purpose was to describe the emergent properties of the data. In phase one, three researchers independently coded the written reports (page three of the booklet). Each coder built up categories of responses that shared thematic causal relationships. Although there was no limit to the number of categories that each coder could generate, the number of identifiable themes that emerged was relatively low. In phase two, a fourth researcher (blind to the raw data) examined each of the three coders' categories searching for common classifications. The result was a coding frame with eight categories of information type and one category for uncodable responses. In consultation with the original coders, precise definitions were generated for each category. In phase two, the three coders learnt the definitions and how to apply them to the original data. In phase four, the original coders used the coding frame to independently recode all of the responses again.

The mean inter-rater agreement was high (median Kappa-Cohen = .84). Less than 1% of responses were judged uncodable. Table 1 shows the categories of causal explanations and the percentage of responses in each category. The modal explanation for differences in quality- of-life measures in countries around the world was wealth.

Table 1. *Percentage of Responses by Causal Explanation.*

	Causal Explanation							
	Wealth	Govt.	Culture	Educ.	Health	Conflict	Pop.	Geog.
% Responses	61	21	6	6	3	2	1	1

These data indicate that most respondents were able to identify the optimal and ecologically valid proxy cue for estimating quality-of-life measures. The second most prevalent response, type of government and government policies, has intuitive appeal but does not appear predictive on closer examination. For example, countries with tax funded health care (e.g., Sweden, Britain, and Canada) report lower infant mortality rates than some countries with free market models (e.g., Pakistan and Bolivia). Nevertheless, some wealthy nations also operate free market models that are also highly successful (e.g., Australia and the U.S.). Government intervention is also confounded by wealth; most emerging economies could not sustain adequate state funded healthcare systems even if such systems were actualized, while wealthy countries appear to choose systems on ideological grounds rather than on absolute performance.[5] A small number of

[5] The difference in absolute infant mortality rate between wealthy countries is generally small, irrespective of the systems that govern healthcare. Policy does impact, however, on whom receives the best treatment. For example, in the sixty largest U.S. cities black children are more than twice as likely to die (14.2/1000) as white children (6/1000). Centers for Disease Control (2002).

explanations were non-predictive (i.e., population) or difficult to operationalize, such as cultural differences.

These data support the Brunswikian view that most people are able to identify wealth as a variable that has an ecologically valid correlation with infant mortality. It also seems that respondents were able to identify the causal relationship between how wealthy people are and the probability of their children dying. Even so, a significant proportion of participants reported causal factors that could be difficult to use consistently across a set of representative items, variables confounded by wealth, or ones that lacked predictive power.

Although the data suggest that proxy variables are potentially an important source of ordinal information (and one that is better understood than infant mortality) these data show that not all respondents will use proxy variables to estimate infant mortality. It also remains to be seen whether those respondents who are able to identify the optimal proxy cue will use the cue for every target item from a set of targets. Experiment 2 extends the initial findings by investigating the cues people refer to when estimating infant mortality rates, and compares respondents' metric and ordinal accuracy. If proxy variables are a source of accurate ordinal information and people are able to generate plausible ordinal inferences from proxy cues, then the metrics and mapping framework predicts that cue usage should be high, and that ordinal accuracy should also be high compared to metric accuracy.

Experiment 2

Method

Participants

Twenty-eight University of Alberta students participated (20 females and 8 males). Participants were Canadian born with a median age of 19 years. Participants received course credit for a procedure lasting approximately 45 minutes.

Materials and Procedure

The stimuli were the twenty countries used in Experiment 1, with one exception. Afghanistan was replaced with Rwanda (infant mortality = 119/1000, pcGDP = Can. $ 1552) due to the war in Afghanistan at that time, which may have disproportionately influenced people's judgments. This study consisted of two tasks. In each, the countries were serially presented and uniquely randomized for each task and each participant. The presentation of stimuli and the collection of numeric responses were computerized, using a standard 17" CRT display and standard 101/102-keyboard array. In Task 1, on the presentation of each country, participants were asked to think about how much they knew about the target country. They then rated each country using a 0-8 scale where 0 = "I have never heard of this country" to 8 = "I know a great deal about this country". Participants were informed that the numbers 0 and 8 represented two extreme ends of a scale and to consider their responses carefully. When satisfied with their judgments, they were told to type a response into an input field using the keyboard's numeric keypad. They could then press the ENTER key to log their response and begin another rating trial. The knowledge-rating task served two purposes. First, it familiarizes participants with the stimulus set prior to estimating infant mortality rates. Second, the percentage of guessed responses, and the overall effect of guessing, can be analyzed if necessary (i.e., "zero" rated countries). This was completed without the presence of the researcher.

At the end of Task 1, the participants were automatically shown a set of instructions. Once these had been read, a researcher checked to ensure the instructions had been understood. In Task 2, participants were told how infant mortality rates are measured, and asked to provide infant mortality estimates for each country. They were also asked to provide concurrent verbal protocols during each trial that would be tape-recorded in the presence of the researcher. On the initiation of each trial, the participant was first required to read the name of the target country aloud for the verbal record. Participants then estimated the country's infant mortality rate and were asked to report their thoughts as they came to mind, to describe how they formulated their judgments, and to include any information that came into consciousness. If respondents fell silent for more than 3-4 seconds the researcher prompted the participant to continue verbalizing, otherwise the researcher sat behind the respondent and remained silent. When the respondent had arrived at an estimate for each country they stated their answer aloud, at the same time as pressing the keyboard's SPACE BAR. This caused an input field to appear, and the participant was required to enter the same response into an input field using the keypad. Pressing ENTER logged each response, and the procedure was repeated with another stimuli. The computer program did not permit non-responses or responses outside of the infant mortality scale.

Results and Discussion

Three analyses were conducted. The first examined the cues that respondents reported, as measured by the type of information frequencies reported in the concurrent verbal protocols. The second analysis compares metric and ordinal accuracy of the infant mortality estimates. Metric accuracy was determined by mean OME calculations for each participant. An ordinal accuracy score was derived for each participant by calculating the rank order correlation between estimated and actual mortality rates. The third analysis was a subsidiary analysis examining the relationship between estimated infant mortality that will be reported in the same section as metric and ordinal accuracy. A cue use score was determined for each participant by calculating the absolute rank order correlation between estimated infant mortality and actual pcGDP. Inferential statistics for cue use were conducted on the Fisher Z-transformed correlation coefficients, and reported means are back-transformed.

Reported Cues

Two researchers independently coded the verbal protocols using the same framework constructed for Experiment 1, with an additional category to include other information, such as personal or anecdotal information (e.g., "A while ago I saw something about this on the TV"). Because the verbal reports consisted of open-ended protocols, respondents were able to refer to multiple sources of information and were free to refer to the same category multiple times. As a result information elicited on a single trial could be coded under more than one category. Multiple references to the same category were scored as "1" as opposed to the absolute frequency, hence the unit of analysis for reported cues is the percentage estimates per category (Max $n = 560$). Less than 1% of reports were uncodable, and inter-rater reliability was high across categories (median Kappa-Cohen = .94).

Again the predominant cue was wealth, with 75% of all estimates referring to wealth at least once. Participants also mentioned other sources of information but less frequently than wealth cues. Table 2 shows the types of information respondents reported, and the percentage of estimates per category. The mean number of categories referred to per estimate was 1.8, although the modal number of cues used per estimate was one (41% of all estimates). Figure 1 shows the mean percentage of responses that can be categorized as using wealth only, non-wealth cues only, both wealth and non-wealth cues, and unjustified responses. Although it appears possible to draw inferences and make infant mortality estimates on the basis of a single proxy cue (30%), respondents were twice as likely to select at least one additional category of information.

Nevertheless, the pattern of cue selection is more complex than suggested by the cue selection frequencies alone, with participants using different types of information, and combinations of information. Figure 2 shows the frequency of each cue selection type

Table 2. *Percentage of All Estimates Eliciting at Least One References as a Function of Category.*

	Type of Information								
	Wealth	Govt.	Culture	Educ.	Health	Conflict	Pop.	Geog.	Other
% of Estimates	75	15	4	2	32	16	11	13	10

for each participant, showing frequencies for use of wealth alone, wealth in combination with other cues, non-wealth cues only, and unjustified responses. These data show a wide range of between participant selection strategies. For every selection choice there is at least one participant who did not employ that strategy for any of their estimates. However, cue selection strategies varied with respondents selected wealth only information up to seventeen times, non-wealth information up to eleven times, both types wealth and other information up to twenty times, with only unjustified responses showing any kind of between subjects agreement (max. frequency < 5). This is not to suggest that information selection is random, or unsystematic. The between subjects variation is likely to reflect differences in both item and domain specific knowledge that each participant was able to retrieve. Still, a significant proportion of responses included references to wealth, supporting the hypothesis that respondents are able to identify and use wealth information when generating numerical estimates.

In addition to the analyzing between subject differences, it also worthwhile examining differences between target items. Because the proxy/ordinal-conversion approach emphasizes item specific knowledge as well as task relevant information as critical to cue selection, different items should evoke different cue selection strategies. The use of other, non-wealth, information did tend to be item specific. For example, across all 560 responses, education was mentioned for only six countries, and Russia's geography was twice as likely to be mentioned than the geography for any other country. Similarly, population was also referenced, but mostly for Japan and India which have large populations. The proxy/ordinal conversion hypothesis predicts that in some cases a single cue will be sufficient to generate an estimate for an item, while for other items proxy cues can be used to determine assignment to a partition but additional cues may be accessed in order to make a final estimate depending on how the proxy cue is able to discriminate between items.

In summary, the cue selection data concur with Experiment 1 by providing more direct evidence that people often generate inferences using ecologically valid proxy variables. Wealth was the most frequent cue referred to, although respondents also

Figure 1. *Percentage of all responses as a function of information referenced.*

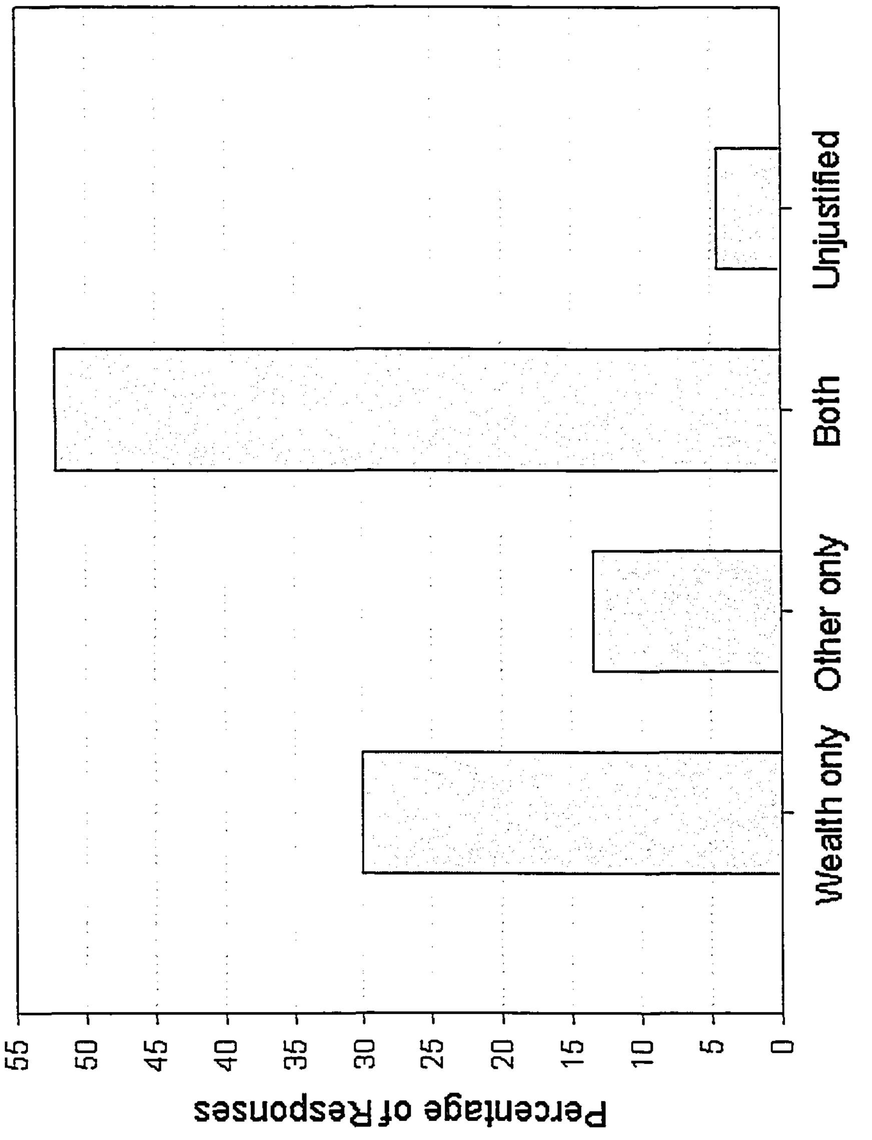
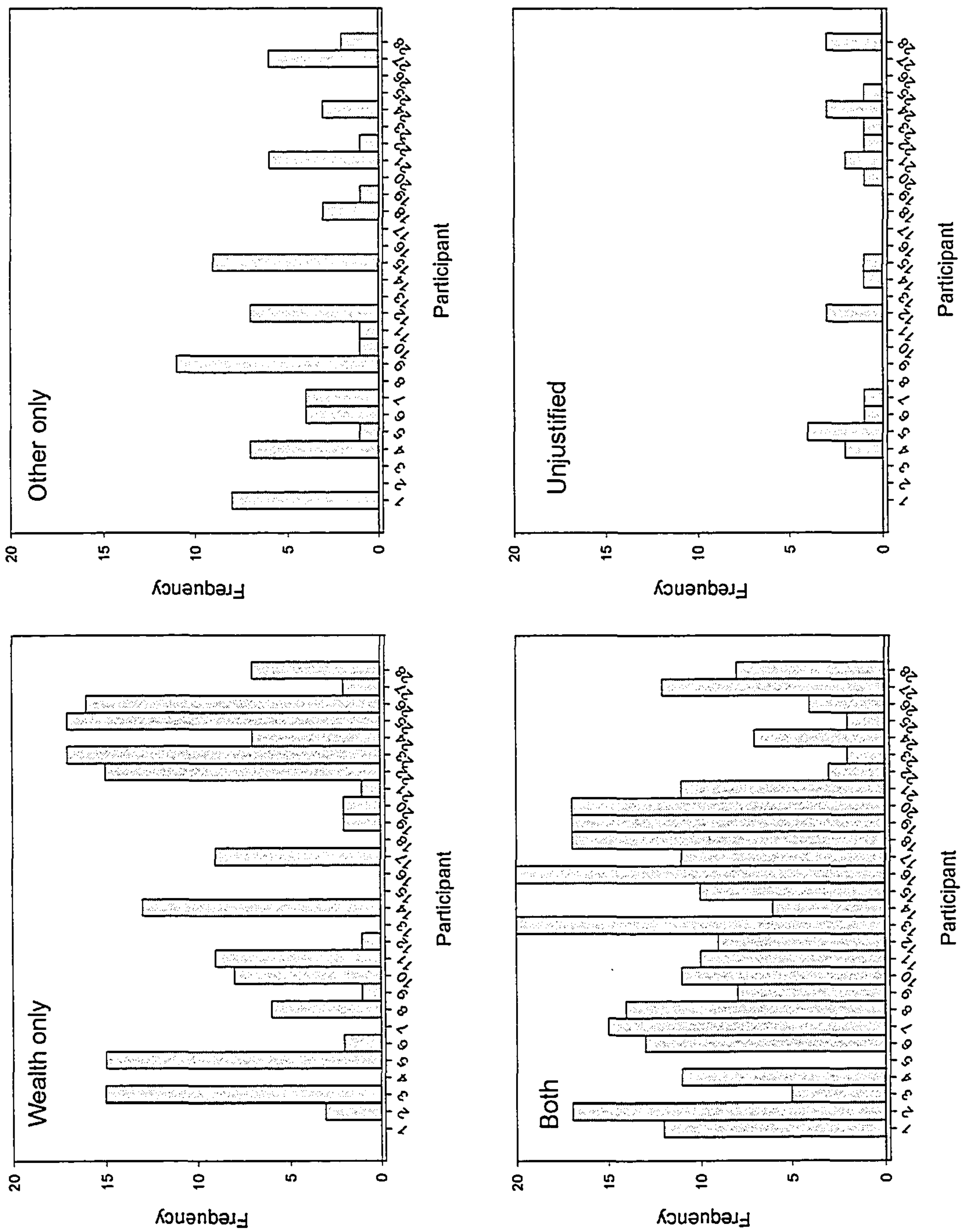


Figure 2. *Frequency of information types referenced by each participant.*



referenced other information. Nonetheless, this additional information tended to be item specific.

Accuracy

The actual mean mortality rate for these countries was 39 deaths per thousand births. The mean metric accuracy across all participants was poor, and over four times the true mean, mean *OME* = .74 (*SEM* = .06). Most participants tended to overestimate (26 participants), supporting the bias reported in Experiment 1. The metrics and mapping perspective stresses that metric accuracy is dependent on domain specific metric knowledge, such as range and central tendency. Again, the obtained OME is consistent with the assumption that respondents have a poor understanding of the target variables metric properties.

However, previous research has consistently shown that metric and ordinal accuracy are most often independent. Because proxy variables provide access to ordinally relevant cues, ordinal-conversion predicts that respondents should be reasonably good at rank ordering items using wealth information, even when their metric understanding is poorly calibrated. An analysis of ordinal accuracy supports this conclusion. Mean rank order correlation between estimated and actual infant mortality was high, $r_s^{map} = .71$ (*SEM* = .04), and significantly greater than zero, $t(27) = 22.8, p < .01$. Since optimal performance using wealth as a proxy variable would result in approximate rank order accuracy of $r_s = .86$, it can be assumed that respondents' subjective intuitions about how wealth impacts infant mortality is quite accurate. This is shown by the fact that estimates correlated more highly with actual pcGDP, $r_s^{gdp} = -.76$ (*SEM* = .05), than actual infant mortality, $t(27) = -3.49, p < .01$. In addition, the relationship between estimated infant mortality and actual pcGDP was in the direction predicted by the assumption that greater wealth is associated with infant survival (all participants' correlations were negative).

Proxy variables were defined as sources of ordinal information that people draw upon when making numerical estimates. Proxy variables were characterized as good predictors of the target domain, were better known than the target domain, perceived as causally related, and understood to predict in a known direction. Experiments 1 and 2 demonstrate that respondents consistently identified wealth as the best predictor of infant mortality and frequently used wealth to generate estimates, indicating that wealth was better understood than the actual mortality rates. First, estimates corresponded more highly with actual wealth than true infant mortality, indicating better ordinal intuitions about these countries financial standings than children's health. Second, respondents were more biased towards overestimation and less accurate compared with participants in the pilot study who estimated pcGDP for the same countries presented in Experiment 2, suggesting that people have less biased and more accurate intuitions about wealth than infant mortality (although the comparison is between experiments, estimates in Experiment 2 correlated more highly with actual pcGDP than actual infant mortality). Participants in Experiment 2 also appeared to have accurate notions about the role of poverty and its effects on children's well being, evidenced by a universal agreement on the direction of the association between poverty and infant death. The evidence indicates that the wealth information people retrieved qualifies as a proxy variable, and it also seems likely that people will use this same information to estimate other quality-of-life statistics such as literacy rates and life expectancies.

These results lend tentative support that participants relied on ordinal-conversion processes, and that these processes were evoked as a result of respondents retrieving proxy information. The discrepancy between metric and ordinal accuracy also concurs with previous findings. First, people's metric understanding for many numerical domains, such as infant mortality statistics, is weak because numerical values are rarely stored (if they were ever learnt). Second, it appears that respondents are able to draw on proxy cues to help them rank order estimates and to make principled estimates based on the information they can retrieve. Although metric accuracy is poor, ordinal-conversion serves to maintain consistency between estimates, and provides a basis for making numerical responses (although other item specific information may also be used). Nonetheless, ordinal inferences cannot be used as a substitute for numerical information, and most respondents failed to grasp an accurate sense of the true range, resulting in gross overestimation.

Chapter III

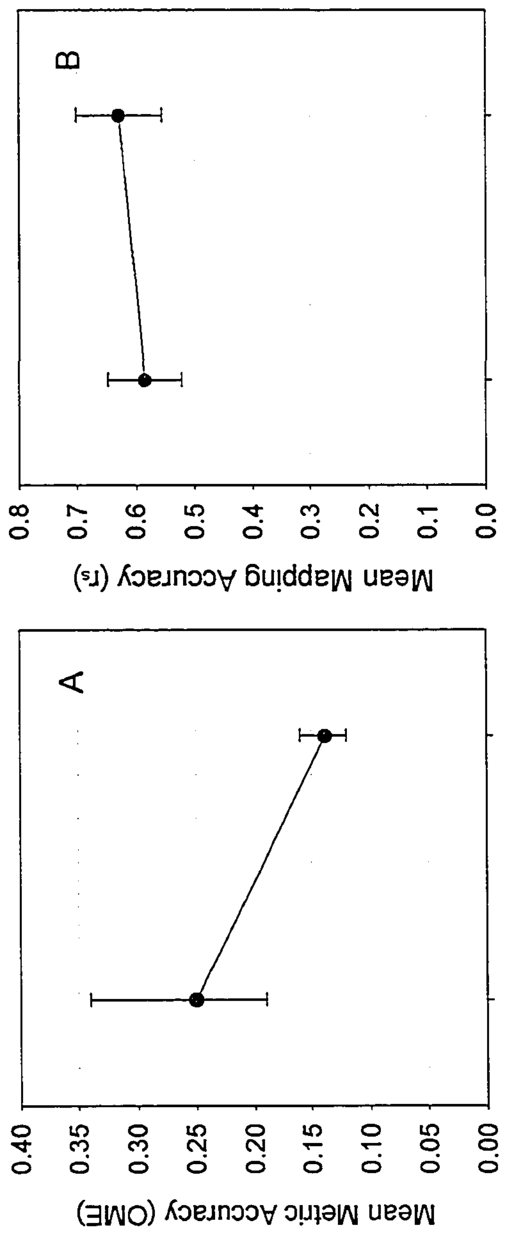
Experiments 1 and 2 provided grounds for characterizing proxy variables as the modal source of ordinal information when estimating infant mortality rates, and possibly quality-of-life measures in general. Experiment 2 also provided tentative evidence for accepting ordinal-conversion as the modal estimation strategy, on the grounds that ordinal accuracy was high despite poor appreciation of the actual metric values. One difficulty with accepting this evidence outright, is that concurrent verbal protocols are liable to capture the content of thoughts, but do not provide an exact description of the estimation processes used. Furthermore, the presence of a researcher, together with the requirement to think aloud, may have resulted in participants deliberating more than they would under real-world conditions. Wilson and Schooler (1991) have pointed out that over deliberation can result in a reduction in performance and preferences that are less optimal than might otherwise be expected (see also, Wilson, et. al., 1993). Still, ordinal-conversion assumes a set of serial processes that are open to introspection and tractable enough to be articulated, as opposed to concurrent parallel processes that can mutually interfere with both decision making and self-reporting by placing too great a burden on working memory (Ericsson & Simon, 1980). More problematic for this research project, is the absence of a method for coding only information that contributed to the estimation process, and eliminating information that may have been considered but disregarded; a scenario that the retrieval-inference cycle predicts when retrieved information fails to produce an inference. For example, wealth appeared to be a modal form of information by most participants across all the target countries, whereas the low frequency categories (e.g., Russia's geography or India's population) may simply be the retrieval of known facts, which may have subsequently been rejected as irrelevant to the task or target item.

It would be imprudent to draw strong conclusions for the proxy/ordinal-conversion hypothesis in the absence of a substantive experimental method capable of isolating the effects of proxy variables from other types of information. Experiment 3 is designed to dissociate ordinal from metric results by controlling respondents' access to accurate information and comparing how proxy cues and metric information have independent effects on performance. The metrics and mapping approach assumes that factors influencing ordinal and metric accuracy are most often independent of one another. In consequence, access to proxy cues should have dissociated effects on people's ordinal accuracy compared with accurate metric information.

The seeding method used to examine how changes in metric knowledge impacts metric accuracy has been well documented (Brown & Siegler, 1993, 2001; Friedman & Brown, 2000b; LaVoie, et. al., 2002; Lawson, & Bhagat, 2002). Ordinal-conversion, on the other hand, has not been the subjected to any principled investigation due, in part, by the absence of an appropriate experimental procedure. The focus of Experiment 3 is to address this imperative by introducing a new methodology called proxy cueing.

Proxy cueing is an evolution of the repeated measures seeding methodology described in Chapter I. In both paradigms participants provide two sets of estimates for the same set of target items. The critical manipulation is the introduction of accurate information during the second round of estimates. In seeding, the newly presented information consists of true values for a subset of items, called seed facts. Figure 3 (unpublished pilot data) shows the characteristic effects of seeding. In this example, participants estimated the literacy rates for forty countries in Estimate 1, and were

Figure 3. *Effect of seed facts on metric and mapping accuracy for estimates of literacy rates.*



* Error bars = 95% confidence intervals (all graphs).

subsequently seeded with the actual rates for a subset of target items in Estimate 2. With seeded items removed from the analysis, these data show that seeding improves metric accuracy for the non-seeded (i.e., transfer) items, Figure 3a. This is consistent with the metrics and mapping approach because seeds convey domain specific metric information. The introduction of accurate metric information allows respondents to recalibrate their metric beliefs. For example, if a participant initially believes that literacy rates are typically about, say, 80% but is subsequently presented with facts indicating most countries have rates higher than 90%, they can infer than their previous estimates were probably too low.

Expressed in terms of ordinal-conversion, seed facts should influence how people set the metric and determine the range (the two initial stages in ordinal-conversion). As a result, seed information has an impact at the domain level, since all of the to-be-estimated items fall within the respondents' subjective notions about the metric and plausible numerical range. Metric beliefs are unlikely to change in the absence of any reason or evidence for doing so, as this is likely to lead to unjustifiable inconsistencies. When people update their metric understanding when presented with seed facts there is a subsequent transfer effect onto non-seeded items in light of new metric understanding. However, this transfer effect has been shown to have little or no impact on ordinal accuracy because seed facts provide no item level information about the relative ordering of target items (Figure 3b).

Nonetheless, the logic of seeding should also apply to ordinal information, insofar that people respond by revising their beliefs when they encounter new facts that they consider accurate and relevant. Instead of respondents receiving domain level information, proxy cueing involves the presentation of accurate proxy information for each target item. Experiments 1 and 2 showed that wealth information is likely to be ordinally relevant when estimating quality-of-life statistics. If ordinal-conversion is the preferred strategy when estimating infant mortality rates, then providing one group of respondents with proxy cues and another group with seed facts during Estimate 2 should result in the following dissociations. First, ordinal accuracy should be moderately high for all of the initial estimates based on the evidence from Experiment 2 that people appear to have fairly accurate intuitions about the fiscal standings between countries at the outset. Still, proxy cueing should increase ordinal accuracy as respondents make use of accurate proxy cues rather than their subjective intuitions, with one critical caveat. Unlike seeds, the predicted effect of proxy cues does not apply if only a subset of proxy cues is provided. Proxy cues are assumed to provide information used to assign targets to partitions, which is a later stage in ordinal-conversion and one that applies at the item level rather than domain level. It follows that a subset of proxy cues should have no ordinal bearing, and thus no transfer effect, on the non-cued items due to a lack of association between the proxy cue for one target item and the infant mortality rate for another. For example, knowing the wealth for one country is not predictive of the relative rank order mortality rate in other countries. However, if proxy cues are provided for all of the target items then an increase in ordinal accuracy should be observed. On the other hand, seed information should have no impact on ordinal accuracy because the information is ordinally uninformative.

By this view, seeding and proxy cueing should also demonstrate a very different pattern of results for metric accuracy. First, metric accuracy should be unaffected by proxy cueing. Experiments 1 & 2 demonstrated that respondents have poor metric intuitions about infant mortality rates. Providing respondents with pcGDP information conveys nothing informative about the magnitude of infant mortality rates. This line of reasoning also applies to a subset of proxy cues. In contrast, seeding is predicted to increase metric accuracy. Therefore a transfer effect can be expected to increase metric accuracy for the transfer items at Estimate 2.

In the following experiment participants were asked to provide two sets of infant mortality estimates for the same target countries. During Estimate 1, all participants made numerical judgments in the absence of any experimenter-given information. At Estimate 2, some participants received seed information; accurate infant mortality information for a representative subset of countries. Two other groups were assigned to receive proxy cues (accurate pcGDP information) for either all the target items or the same subset of representative countries used for the seed group. Finally, a control group re-estimated the target items again but were given no additional information.

In addition to the impact of accuracy, differences in cue usage (as measured by the rank order correlation between estimated infant mortality and pcGDP) should be observed. If people respond to proxy information in the way that the proxy/ordinal-conversion hypothesis assumes, then the correlation between estimates and proxy information should increase; such that during Estimate 1, participants are believed to be using their subjective understanding about the wealth of the countries they are asked to estimate. Participants who receive proxy cues during Estimate 2 for all of the items should use this information by mapping their responses onto the proxy variable.

Experiment 3

Method

Unless specified, methodological features common with Experiment 2 are not repeated.

Participants

One hundred and sixty undergraduate students (110 females & 50 males) participated in a procedure lasting approximately one hour. Median age was 19.

Materials and Procedure

Fifty-eight representative countries were selected as stimuli (Appendix II). The experiment consisted of three computer-controlled tasks. In Task 1, each country was presented in turn and participants rated their knowledge of each using the 0-8 point scale. In Task 2 asked participants to provide initial infant mortality estimates for the same set of countries. In Task 3, participants were randomly assigned into four equally sized groups that determined the type and quantity of information they would receive when estimating the countries for a second time. These were either; a) seed facts for a subset of seven countries (the actual infant mortality rates for: New Zealand, Cuba, United States, Poland, Nicaragua, Nepal, and Afghanistan); b) a subset of proxy cues (the pcGDP for the same seven countries as the seed group); c) pcGDP proxy cues for all 58 countries; or d) a control condition where participants received no information, but were still required to estimate the countries for a second time.

Before Task 3, all participants were instructed that they would be estimating the mortality rates for the same countries again. In the seed and proxy cued conditions the relevant information was described in detail. For example, it was explained that per capita Gross Domestic Product was a measure of national wealth, controlling for population size. Those who received either seeds or the subset of proxy cues were also told that the information for all seven countries was accurate and would be available on-screen throughout Task 3. Participants who received all of the proxy-cues were informed that the pcGDP information was accurate and would be displayed on a per item basis; when each target country was presented its corresponding pcGDP also appeared on screen, and remained visible until the respondent had completed their estimate. All participants who received information were told that while this information was being made available they were not obliged to use it, and that they may or may not find the information relevant. All participants completed the experiment alone in a quiet laboratory with the door closed.

Results and Discussion

Unless specified, the seven countries shown as either seeds or proxy subsets were eliminated from the following data analyses. Two analyses surveyed metric and ordinal accuracy between groups that received different types and different quantities of information and the final analysis examined cue usage.

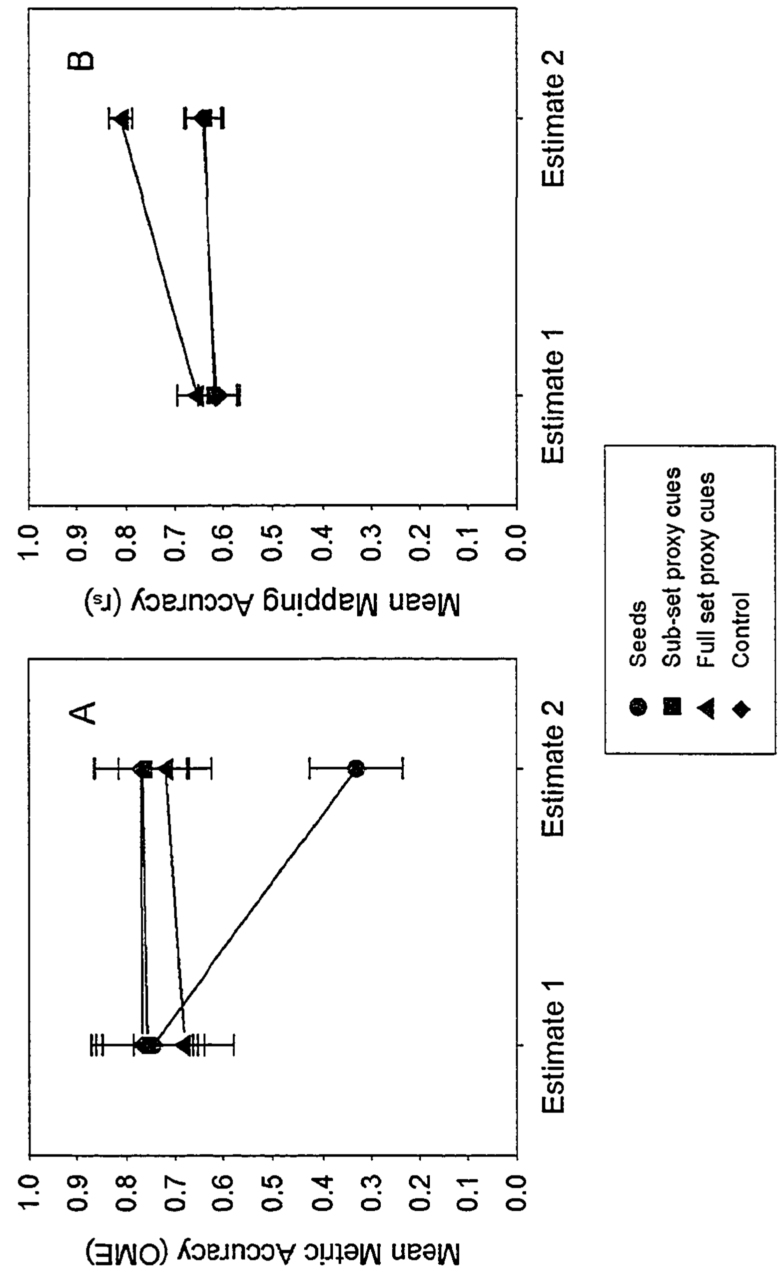
Metric Accuracy

Mean OME figures for Estimates 1 and 2 were calculated for each participant. Collapsing across conditions, metric accuracy for the initial estimates was poor and equivalent to that found in Experiment 3, mean $OME = .74$ ($SEM = .03$). No differences between groups were expected or found at Estimate 1, $p > .05$. To determine the effect of information type on changes in metric accuracy between Estimates 1 and 2 a 2 (estimate) x 4 (group) mixed factor ANOVA was conducted on mean OME. Figure 4a illustrates a significant interaction between Estimates 1 and 2, $F(3, 156) = 52.0$, $MSE = .018$, $p < .01$, as a function of each information group. A test for simple effects showed that seed facts were the only information type to affect metric accuracy. These data concur with previous studies demonstrating that seeds provide accurate numerical information allowing respondents to update their understanding of the numerical properties of the domain, and thus improve their metric accuracy. As in Experiment 2 respondents tended to overestimate initially, but adjusted their estimates downwards when presented with seed information. Because pcGDP information does not imply anything about the magnitude of infant mortality rates neither of the two proxy groups nor the controls made significant adjustments in the magnitude of their estimates, indicating that wealth information had no impact on these participants metric understanding about infant mortality rates. The metrics and mapping framework accounts for this dissociation because wealth information is used to make reasoned judgments about the relative ordering of countries, not the magnitude of infant mortality rates.

Ordinal Accuracy

Rank order correlations between estimated and actual mortality rates were calculated on first and second estimates for each participant. Initial estimates were again relatively accurate, $r_s^{Est1} = .63$ (*SEM* = .02). As expected there were no significant differences between groups at Estimate 1. Figure 4a shows a significant increase in ordinal accuracy when participants received the full set of proxy cues, $r_s^{Est2} = .81$ (*SEM* = .03), but neither the seeds nor the subset of proxy cues had any effect on ordinal accuracy. A 2 (estimate) x 4 (group) mixed factor ANOVA confirmed this interaction, $F(3, 156) = 25.61$, $MSE = .018$, $p < .01$. Multiple comparisons confirmed that presenting respondents with wealth information for each country increased mapping accuracy, but there was no significant change in any other groups. A summary of metric and ordinal accuracy

Figure 4. *Metric and mapping accuracy as a function of cue group.*



illustrates an important dissociation between seed and cue types on different accuracy measures. These data confirm that proxy cueing is a technique that can be used to explore ordinal-conversion; providing access to accurate ordinal information for all items impacts ordinal accuracy but has no effect on metric accuracy. Similarly, accurate metric information influences metric but not ordinal accuracy. This double dissociation presents additional evidence for the independence between different types of information and measures of accuracy as proposed by Brown and Siegler (1993).

The proxy/ordinal-conversion hypothesis predicted that the proxy cueing manipulation should increase ordinal accuracy when proxy cues were shown for all countries, while seeds should affect metric accuracy. This delineation can be drawn because ordinal-conversion is an estimation strategy particularly useful in domains where metric knowledge is sparse. Experiment 3 is a demonstration of ordinal-conversion's multi-process architecture. Domain level processes determine the range and metric properties that affect all estimates (at least until a sufficient reason is found to revise these metric beliefs, i.e., seed information). Item level processes, on the other hand, are used to assign numerical values to individual target items. It follows that a subset of seed information was sufficient to evoke changes in metric beliefs at the domain level, although a subset of proxy cues for the same items had no effect on ordinal accuracy. As previously demonstrated, respondents are able to identify cues that are good ordinal predictors and at perceiving a causal relationship between the proxy and criterion variables. It appears that the presentation of proxy cues for all target items serves to augment the respondents' existing ability to discriminate between the countries on the basis of wealth, with a corresponding increase in ordinal accuracy due to the predictive power of wealth as a proxy variable.

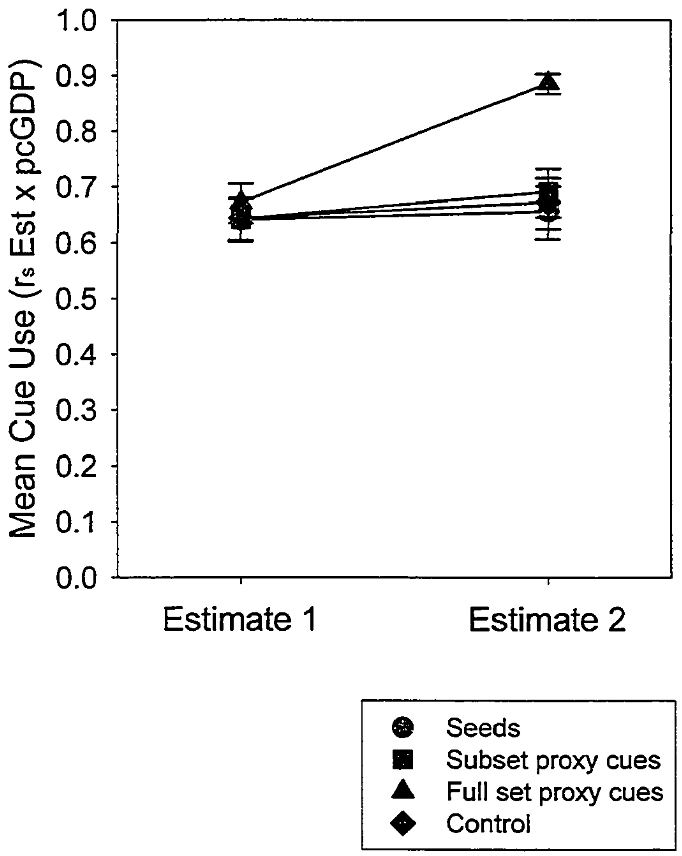
These results also provide insight into how respondents make use of proxy information. The absence of a transfer effect for the pcGDP subset group also indicates that proxy cues do not change people's beliefs at the domain level. There is no evidence, for example, that respondents evoke and reinterpret the relationship between wealth and infant mortality. Analyzing the countries used as proxy cues is particularly informative. Paired comparisons showed that respondents who received proxy cues for all countries also showed an increase in accuracy for both the seven subset countries, $t(39) = -2.22, p < .05$, as well as the remaining fifty-one items, $t(39) = -8.97, p < .01$. Tellingly, those who received a subset of proxy cues were no more accurate at rank ordering the seven subset countries, $t(39) = -1.07, p > .1$, or better at ordering the remaining non-cued items, $t(39) = -1.55, p > .1$. Despite the fact that the same wealth information was available for the same seven items for both groups, the subset group appear to have ignored the information because it was only relevant 12% of the time, suggesting that participants are aware that proxy information has little value if it cannot be applied across all items. An alternative explanation is that participants were able to maintain high rank order accuracy using subjective information alone. Perhaps a subset of accurate information would seemingly disturb the overall relationship between subjective knowledge and estimation strategy by introducing a few accurate figures that (if taken into account) are likely to destabilize relationships within the set as a whole. This interpretation is supported by the cue use correlations. The subset group did not increase their rank ordering for the seven cued countries, $t(39) = -.38, p > .1$, compared with an increase in cue use for the seven subset countries shown by the full set proxy group, $t(39) = -5.9, p < .01$. This is informative because it suggests that people have subtle ideas about the relative merits of information quantity, and may be capable of distinguishing between information types, and the quantity of information required before it can be used to generate plausible estimates.

In summary, the proxy/ordinal-conversion position predicted that transfer effects only occur if respondents change their beliefs at the domain level, because domain level information has impact across all items. On the other hand, the fact that the full set of proxy cues had a large effect on ordinal accuracy shows that proxy variables work at the item level. Knowing the wealth of a country has direct ramifications for that country's infant mortality rate, but little relevance to the infant mortality rates of the other nations.

Cue Usage

pcGDP cue use, determined by the absolute rank order correlation between estimated mortality and pcGDP, was predicted to support the accuracy findings. Figure 4b reveals that initial estimates correlated well with actual pcGDP, $r_s^{gdp1} = .65$ ($SEM = .06$), and that cue use increased when the full set of proxy cues was presented, $r_s^{gdp2} = .89$ ($SEM = .04$). There was no change in pcGDP cue use for the proxy cue subset group, and no cue pcGDP cue use for either the seed or control groups (since pcGDP information was not made available). Figure 5 shows the pattern of cue use by each information group, with the interaction confirmed by a 2 (estimate) x 4 (group) mixed factor ANOVA applied to the cue use correlations, $F(3, 156) = 54.08$, $MSE = .027$, $p < .01$. A test for simple effects showed changes in cue use for the full set group only. For this group, access to accurate wealth information for each item accounts for their increase in ordinal accuracy as respondents appear to have adopted this information on the understanding that wealth is likely to predict the probability of children surviving. Conversely, a subset

Figure 5. *pcGDP cue use by cue group.*



of proxy cues had no effect on non-cued items. This is probably because there is no obvious relationship between the wealth of different nations, or transparent causal relationship between the wealth of one nation and the infant mortality of another. On this basis, it appears that the subset of proxy cues was rejected in favour of the participant's own subjective beliefs, which provided adequate (if not optimal) performance during both sets of estimates.

However, a criticism of this interpretation is the possibility that experimental demand accounts for the change in cue use. Participants may be motivated to rank order their estimates using wealth on the grounds that the researcher provided the information, not because the cues were considered ordinally informative. The data do not indicate if respondents used the same cognitive strategies to generate both their estimates, or if they gave up making principled judgments by blindly matching their responses onto the cue information. The evidence from Experiments 1 and 2 demonstrated that people understand the validity of wealth in reference to quality-of-life statistics, and make liberal references to wealth information when estimating infant mortality rates. Having made estimates that correlated more highly with actual pcGDP than actual infant mortality during Estimate 1, it seems unlikely that the proxy cues were treated in an arbitrary fashion. Furthermore, no demand effect was observed for the seven items for the subset-proxy cue group, despite the fact that the pcGDP information being also openly available.

The relevant conclusions that can be drawn from Experiment 3 are twofold. First, the proxy cueing method was shown to affect ordinal accuracy without impacting on metric accuracy. In addition, the type and quantity of cues are consistent with the proxy/ordinal-conversion position, insofar that proxy information serves at the item level after a range and partitions have been decided. Second, there is evidence that people are able to evaluate and reject cues that lack the necessary quantities needed to generate ordinally ranked estimates.

In the following two chapters, the proxy/ordinal-conversion hypothesis is tested under increasingly robust experimental conditions. In addition, Experiment 4 compares the demand effects argument against a cue evaluation and selection position by exploring the notion that people are able to identify good from bad predictors, and reject the latter as ordinally uninformative. Evaluation and selection on the basis of ecological validity is the antithesis of experimenter demands (which should lead to cue acceptance, not rejection).

Chapter IV

It was posited that the proxy cueing effect might be due to respondents interpreting the information non-normatively because of situational and experimenter demands placed on them by the procedure. Orne (1962) has pointed out that participants may perceive the experimenter as particularly knowledgeable and the experimental instructions as a prescription to behave in ways that would not normally occur outside of the laboratory. Like Orne's participants, (who tried to bias his results in favour of their own notions about his hypothesis) the participants in Experiment 3 could have perceived the proxy cues as a hint to help prove one theory over another. Alternatively, the experimenter might be viewed as an authority on infant mortality and its causes. Rather than trying to bias the study, participants may have been felt obliged to rank their estimates with the wealth information, as opposed to treating the proxy cues as ordinally informative. At face value, this explanation would appear to invalidate both the proxy cueing method and the conclusions drawn from Experiment 3. However, a closer examination reveals two reasons for arguing against the unconditional matching of estimates onto cues.

First, participants did not map their responses onto a subset of proxy cues any more than the control group who received no information. The absence of a partial demand effect for the subset argues against unconditional acceptance of experimenter provided cues, at least as a general phenomenon. More critically, the demand effect hypothesis is a psychologically implausible explanation for the change in accuracy for the full set group in Experiment 3; all forty participants in the full set group returned negative cue use correlations, indicating that participants not only treated the cues as informative but also understood both the causal relationship between wealth and infant health and the direction of the relationship.

On the other hand, the proxy/ordinal-conversion view implies that a high degree of ordinal accuracy can be achieved by recognizing the likelihood of a high ecological validity between proxy and target variables. Assuming that the target items are representative provides a sound basis for making principled estimates on an item-by-item basis using ecologically valid information. Ordinal-conversion predicts that the only information to be maintained across a set of estimates is range and partition information, and a belief in the predictive value of the proxy variable. This is similar to the QuickEst claim that human rationality can overcome many computational constraints by "letting the environment do the work" (Hertwig, et al., 1999).

Experiment 4 tests these assumptions by manipulating the validity of the cue information. The proxy/ordinal-conversion and demand hypotheses offer different predictions about how good and poor predictors will impact ordinal accuracy. In the following study the target countries were selected to represent the relationship between wealth, population and infant mortality for countries globally. This is to say that the internal validity of pcGDP as a proxy cue is equivalent to the ecological validity (i.e., the predictive power of wealth for all countries always remains high, and the selected stimuli are representative of this relationship). With a representative sample of target countries the ecological validity of population is extremely low (r_s = .02) and is more or less matched by the selected stimuli's internal validity (r_s = .06). In the following experiment participants estimated infant mortality rates, using the same general procedure as full proxy cue group in Experiment 3. Prior to making their second estimates, participants were randomly assigned to one of three groups. These groups were provided with either pcGDP, population, or both types of information for each target item. Experiment 1 showed that only 1% of respondents considered population the primary predictor of infant mortality. In Experiment 2, 11% of responses included references to population, although it is questionable whether this information played a significant role in either partitioning the range or assigning numerical mortality rates to individual targets.

The demand hypothesis predicts that there should be a significant increase in cue usage between first and second estimates irrespective of cue type, and a likely increase for one cue over the other when pcGDP is presented simultaneously with population (assuming that one cue exhibits a greater demand effect over the other). As a result, the demand hypothesis predicts there should be differential effects on ordinal accuracy due to large differences in the predictive power between wealth and population. Experiments 2 and 3 demonstrated that Estimate 1 ordinal accuracy should be fairly high. Still, the demand hypothesis position predicts a moderate increase in ordinal accuracy when pcGDP cues are presented alone (as evidenced in Experiment 3), but a large decrease in accuracy for the population only group since population is an extremely poor predictor of infant mortality. In both of these conditions cue use is predicted to be high for Estimate 2 due to experimenter demands.

The proxy/ordinal-conversion position predicts a very different pattern of results for both the cue use and accuracy correlations. A critical assumption that runs through this research is that people are able to identify proxy cues, generate inferences, and draw conclusions that permit them to rank order their estimates effectively when target items are representatively sampled. The proxy/ordinal-conversion position argues against the unconditional use of cues or that the presence of the experimenter provides cues outweigh people's own intuitions about infant survival. This implies that respondents are capable of evaluating cue information and may decide to accept or reject information according to their own understanding of the numerical estimation problem. When single cues are presented pcGDP but not population should be accepted as useful, resulting in an increase in cue use and accuracy correlations for the former, but no significant change in either correlation for the latter.

If the proxy/ordinal-conversion hypothesis is valid, then cue evaluation and rejection should be most apparent when both cues are presented. Two cues, both of which are accurate but only one of which is predictive, should provide an interesting exploration into how people choose information when generating numerical estimates. The proxy/ordinal-conversion position predicts that performance under dual cues should be equivalent to the pcGDP only situation, assuming that respondents evaluate and reject cues with poor predictive power while utilizing good cues. This prediction has significant theoretical implications. One reason is the general consensus that people access a fairly limited amount of information and often rely on plausible inferences when engaged in a wide range of tasks such as making choices, ordering preferences, and evaluating features, (e.g., Fasolo, McClelland, & Todd, in press; Newman, 1977; Urbany, Dickson, & Wilkie, 1989), but little overall consensus about whether people integrate cue information or rely on single but ecologically valid cues.

It was previously stated that ordinal-conversion was less concerned with the number of cues people use and more focussed on how modal types of information combine with estimation processes to convert ordinal inferences into numerical estimates.

This focus notwithstanding, the approach accepts that more than one cue might be accessed. Experiment 2 showed 41% of estimates were based on references to a single item of information even though the majority of answers mentioned between two and four sources of information. Models based on the principles of the "fast and frugal" approach (Gigerenzer & Goldstein, 1996; Gigerenzer, et. al., 1999) also rely on the principle that information is typically limited or unavailable, but also suggest that respondents make decisions on the basis of single cues alone. Providing respondents with two cues, with very different predictive powers, provides one test of cue selection.

There are four possible outcomes in the dual cue decision space - accept both cues, reject pcGDP and accept population, accept pcGDP and reject population, or reject both cues. A decision to accept one cue and reject another posits a decision mechanism based on the respondents' subjective evaluation of each cue, presumably on the basis of each cues predictive power and the degree to which the cue can discriminate between target items. It would appear reasonable for respondents to accept the most predictive cue, leaving only one viable alternative in the problem space; accept the wealth information, and reject population.

Experiment 4

Method

Participants

Seventy-two undergraduate students participated (48 female and 24 male). Median age was 19. Participants received course credit for a procedure lasting approximately one hour.

Materials and Procedure

Fifty-eight representative countries were selected as targets (Appendix III). The experiment consisted of the three-phase proxy cueing method. In Task 1, participants rated their knowledge of each country. In Task 2 all participants provided an initial set of infant mortality estimates for the same set of countries. Finally, in Task 3, participants were randomly assigned to one of three cue groups, and were asked to estimate the infant mortality rates for a second time. Participants received either pcGDP only information for each country, the population only of each country (to the nearest one hundred thousand people), or both the pcGDP and population figures. When both cues were present, the on-screen information was counterbalanced so that population appeared above the pcGDP information in half the trials. Because Experiment 3 showed that there were no changes in cue use or accuracy when cue information was withheld, no control condition was included. In each condition a precise description of the cue(s) was provided to ensure that participants understood what the numerical information represented and the metrics used to measure them. Participants were told that the information for each country would be freely available when making their estimates, but they were not obliged to use the information if they did not want to.

Results and Discussion

Because the design of this experiment produced a matrix of predictions linking cue type, ordinal accuracy and cue usage, the accuracy and cue use analyses will be presented together broken out by cue condition. Ordinal accuracy was high to begin with, mean $r_s^{Est1} = .62$, and concurs with each of the previous two experiments that people are fairly good at rank ordering infant mortality estimates for a representative sample of countries based on their existing understanding about infant health. A check on random assignment showed no differences in accuracy or subjective cue use between any of the groups at Estimate 1.

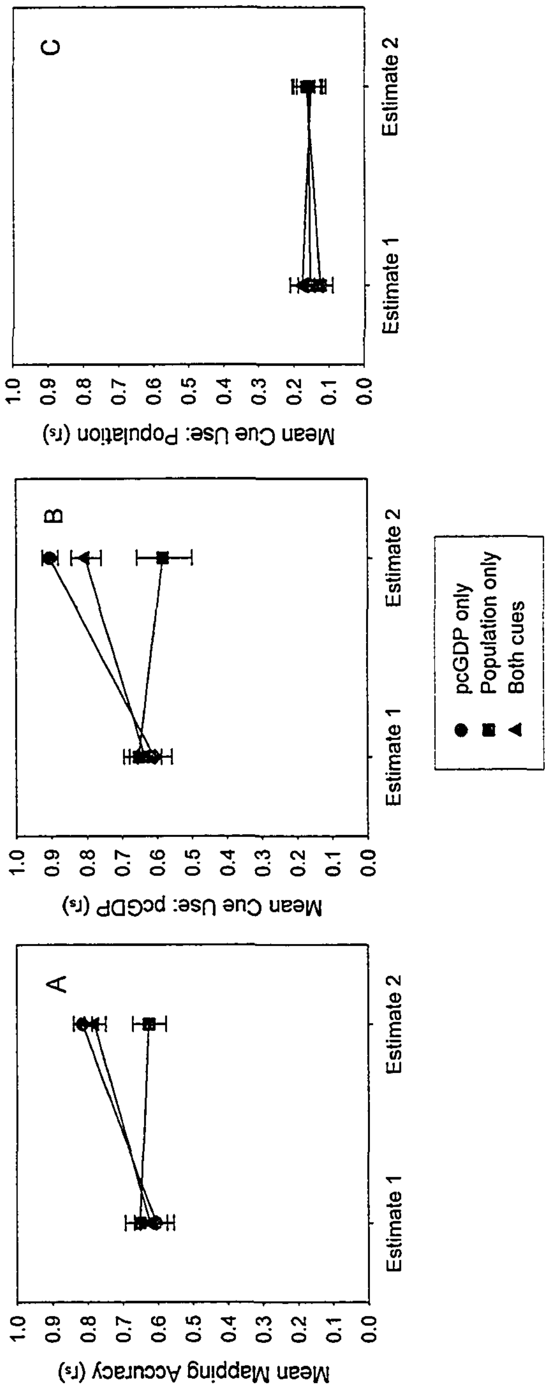
pcGDP Only

Respondents who were provided with pcGDP only information at Estimate 2 showed a significant increase in accuracy, $t(23) = -8.32, p < .01$ (mean ordinal accuracy for Estimate 1 and 2 across all cue groups are illustrated in Figure 6a). Figure 6b illustrates the correlation between estimated mortality and actual pcGDP for Estimates 1 and 2 conditionalized by cue group. For the pcGDP only group, cue use for pcGDP increased from $r_s^{gdp1} = .61$ to $r_s^{gdp2} = .91$, $t(23) = -10.54, p < .01$. Both the demand and proxy/ordinal-conversion hypotheses predicted this result, although for very different reasons. The demand hypothesis suggested that an increase in cue use would be due to participants mapping their estimates onto the pcGDP information because of a desire to help the researcher prove a particular theory or because the cues were hints provided by experimenters who were also perceived as experts. However, observing the effect of population cues on accuracy and cue use can provide a more decisive comparison.

Population Only

No difference in ordinal accuracy between Estimates 1 and 2 was detected for the population only group, $t(23) = 1.06, p > .1$. More importantly, Figure 6c shows that estimates correlated poorly with the actual populations for the initial estimates, $r_s^{pop1} = .13$, and did not alter even when the actual populations were provided for each country, $r_s^{pop2} = .16$, $t(23) = -1.23, p > .1$. Moreover, these coefficients are most likely inflated because of the covariance between actual pcGDP and actual population, $r_s = .17$, and the use of the absolute correlation coefficient between estimate and cue (which overlooks

Figure 6. *Accuracy and cue use as a function of cue group.*



signed differences that may cancel each other out, given that participants are also likely to have differing intuitions about the direction of the relationship between weak predictors, such as population, and infant mortality. When pcGDP is partialed out of the cue use correlations, and signed correlations are permitted, the unique effect of population on cue use is less than $r_s^{pop} < 0.05$ in all cue groups and across both Estimates 1 and 2.

A demand effect explanation for cue use, and associated effects on accuracy, predicts that respondents should have taken notice of experimenter provided information, irrespective of its validity. The population only group showed no evidence that accurate population cues had any impact on their estimates, leaving two reasonable conclusions; participants either rejected the population information or they used the information but were unable to maintain the mapping between the cues and their responses. The later explanation can be tentatively rejected because the pcGDP only group had no difficulty mapping estimates onto pcGDP figures. Given that the numbers describing pcGDP are between four and five digits, compared with the two or three for populations, the population only group should have found the mapping task easier rather than harder than the pcGDP only group due to less demand keeping smaller numbers in memory. The most plausible explanation, then, is that participants are capable of evaluating and rejecting information with poor predictive power, preferring to rely on their existing intuitions about wealth and infant health. In fact, these participants were able to maintain moderately high rank order accuracy for both Estimates 1 and 2, despite the presence of distracting information with potentially devastating effects on their ordinal accuracy.

The finding that ordinal accuracy was maintained between Estimates 1 and 2 provides further insight into how participants treated the population information.

Although these respondents appeared to reject population as a valid source of ordinal information, this does not preclude the possibility that some respondents were perturbed by its presence. Participant's initial estimates correlated with actual pcGDP at r_s^{gdp1} = .65, followed by a small but significant decrease in correspondence with wealth at Estimate 2, $r_s^{gdp2} = .58$, $t(23) = 2.99$, $p < .01$. This result is interesting, given that no differences were found in either ordinal accuracy or the use of the population cues. A subject based analysis sheds partial light on what may have occurred. Two thirds of the participants (n = 16) showed a mean decrease in correspondence of $r_s^{gdp2} = .20$, while the other third showed a far more modest increase, mean increase $r_s^{gdp2} = .08$. It appears that the aggregated change in the strength of the relationship was significant, but insufficient to impact ordinal accuracy. Why this should have occurred is obscure. Perhaps the majority of these participants became distracted from using subjective wealth by the presence of population information and the requirement to evaluate it, despite the fact they were not prepared to use population.

Still, the overall results are consistent with the general point being made; people are able to identify and use predictive cues, while disregarding poor ones. The demand hypothesis does not predict this dissociation, and the evidence indicates that respondents were not simply mapping estimates onto cues because they are available, but because the predictive cues supported pre-existing notions about the true factors that determine infant health around the world.

Both-Cues

The both-cues group should show the same pattern of responses as the pcGDP only group, presupposing that respondents are capable of evaluating cue validity and are able to reject poorly predictive information. This interpretation is supported by the data. Figure 6a shows that ordinal accuracy increased between initial and final estimates, $t(23) = 8.39, p < .01$. This was matched by an increase in pcGDP cue use, Figure 6b, $t(23) = -6.77, p < .01$, but no change in population use was detected, $p > .1$. The pattern of results in Figure 6 were confirmed by a 2 (estimate) x 3 (cue group) mixed factor ANOVA conducted on accuracy, pcGDP use and population use. A significant interaction for the accuracy data, $F(3, 69) = 32.55$, $MSE = .023$, $p < .01$, highlights the increased accuracy shown by the pcGDP only and both cues groups. A significant interaction for pcGDP use, $F(3, 69) = 61.83$, $MSE = .039$, $p < .01$, in the context of a null effect for population use, $F(3, 69) = 1.57$, $MSE = .008$, $p > .1$, serves to show how pcGDP was the only information to influence ordinal accuracy, irrespective of presence of uninformative information.

Experiment 4 provides a more compelling, within subjects, test that people evaluate cue validity and are capable of selecting good from poor cues. In order to gauge the magnitude of the cue effects the absolute differences were calculated for each participant across three measures; ordinal accuracy between Estimates 1 and 2, pcGDP use, and population use. These derived scores provide one criterion variable (absolute change in accuracy) and two predictors (the absolute changes in pcGDP and population usage), on which a hierarchical multiple regression was performed. When pcGDP was entered as a single predictor, this alone accounted for 66% of the variance in accuracy. The addition of population added just 1% of explained variance, controlling for the pcGDP x population covariance.

It was pointed out that ordinal-conversion is a framework whose primary function is to explain how ordinal information combines with estimation processes, and how the information is processed to generate ordinally coherent numerical estimates. In summary, three broad conclusions can be drawn from Experiment 4 agree with the processes assumed by the framework. First, the proxy cueing effects in Experiments 3 and 4 cannot be accounted for by demand characteristics alone. Population made no impact on people's accuracy or the propensity to use wealth as a primary source of ordinal information. Even when actual pcGDP was absent, respondents' estimates correlated more highly with actual pcGDP than actual infant mortality. This replicates the effect found in Experiment 3. When both types of information were freely available, the overwhelming response mode was to map responses onto the predictive cue, and ignore the irrelevant information.

The second conclusion is that people responded to information from the functional standpoint of generating task relevant inferences about each target item. It appears that people understand not only the task, but also the information that best serves the processes used to complete it. Taken together Experiments 3 and 4 show that respondents have a fairly robust subjective understanding of the wealth of the target countries and the relationship between wealth and infant mortality. When accurate wealth information became available, respondents were able to exploit the accuracy of the information, resulting in significant and meaningful increases in ordinal accuracy.

Finally, for the both cues group there were four possible choices in the cue selection problem space. However, there are different explanations for why respondents might be able to select single cues from multiple alternatives, and interesting reasons why some experimental designs could obscure the actual decision process rather than shedding light on them. For example, the fast and frugal approach predicts that only one cue will be used and that the ecological validity is used to determine cue choice. The data from the both-cue group supports this conclusion. However, the idea that the environment can be made to "do the work" (Gigerenzer, et al., 1999; Hertwig, et al., 1999) may be confounded by the types of task used to endorse these claims. To understand why, it becomes necessary to examine the critical assumptions behind "fast and frugal" heuristics, especially the studies used to support the theory more closely. One-reason decision-making assumes that the to-be-estimated items are necessarily representative of the whole population, and that the numerical domain being estimated can be predicted by a single cue. In consequence, the strong claims for one-reason decision models have been made based on limited empirical data from studies that have tended to use methods where single predictive cues are most useful or accessible. Indeed, this criticism can be levelled at Experiments 3 and 4. Replications that are insufficiently differentiated serve, at best, to limit the generalizability of the theory under test (Lindsay & Ehrenberg, 1993), and may, at worst, exclude the possibility that the theory is false (Popper, 1934). If small but meaningful variations in test conditions result in markedly different patterns of results there are strong reasons for doubting the validity of the theory, or at the very least demands a theoretical revision.

It is important, then, to question how different models can account for the data under a range of different conditions. For example, when no single ecologically valid cues exist, or when sets of items are unrepresentative (these issues are not mutually exclusive as cues with high ecologically validity across all items may not demonstrate high internal validity for a subset of items). QuickEst processes remain inflexible by allowing ecological validity to govern performance under all conditions. Proxy/ordinal-conversion is a more flexible approach, arguing that cues may also be chosen on the basis of how relevant they are to the to-be-estimated stimuli (i.e., internal validity), as well as the ecological validity. However, both theories are capable of explaining the findings in Experiment 4, since the stimuli were representatively sampled, and the internal validity of both pcGDP and population matched the ecological validity. The following study attacks these two assertions by testing the degree to which people rely on the environmental structure when the to-be-estimates items are unrepresentative.

This challenges the idea that ecological cue validity is the sole determinant of cue selection. The main questions are whether multiple cues can be used, and whether they are processed in parallel or serially (at different levels in a multiple process strategy). At one level the proxy/ordinal-conversion position concurs with the latter hypothesis. It is also assumed that respondents may interrogate other cues, especially if the information is readily accessible and seems relevant. Here, it is crucially important to treat cue relevance as a task and item dependent characteristic, and not solely on the basis of predictive power across representative samples. For example, Sierra Leone, Malawi and the Democratic Republic of Congo are among the five poorest nations on earth, and are likely to be judged by people as equally poor. Despite our participant's ability to identify wealth as a proxy cue, it is highly unlikely they would be able to discriminate between the $117 pcGDP that separates the poorest, Sierra Leone (pcGDP = $556), from the richest, Congo (pcGDP = $673). If asked to estimate the infant mortality rate for these countries, respondents have ample reasons for assigning all three countries to the high mortality partition, but scant reasons choosing a specific infant mortality rate that differs from the others on the basis of a country's poverty (at this level of poverty, other factors may become the better discriminator of infant mortality rates, i.e., war, famine, and drought). The relevance of ecologically valid cues with high predictive power is altered because a lack of subjective discriminability on the proxy cue may evoke the retrieval of other information. As a result cues that are otherwise non-predictive and lack ecological validity may be used so that countries can still be ordered on principled grounds (presupposing that information can be used at different levels in the ordinal-conversion processes). For example, judging Sierra Leone and the Democratic Republic of Congo as having higher infant mortality rates than Malawi, on the basis that the former are currently, or have recently, been at war, while Malawi has not.

This is to claim that wealth may still be used to assign a target, yet cues that may be unreliable at the general level could still be used to discriminate between targets that are closely matched on the proxy variable. All things being equal, it is reasonable to assume that these otherwise irrelevant facts are either locally predictive, or may be perceived by respondents as locally predictive. Experiment 5 provides a replication of the paradigm used in Experiment 4, but manipulates the internal validity of the cues. This approach provides a major test of the underlying theoretical assumptions that have underpinned the studies presented so far, and demands that the proxy/ordinal-conversion model of decision making describe how multiple cues can be used to make numerical decisions.

Chapter V

Experiment 4 concluded by making three assertions. The primary contention was that the methods used to test one-reason decision-making might underestimate the complexity of how information is structured in the real world. For example, the *recognition heuristic* (Goldstein & Gigerenzer, 1998) suggests that binary decisions, such as comparing two city populations, can be accurately judged when one city is recognized and the other is not (on the basis that recognized cities are likely to be larger than unrecognized ones). While plausible in the context of city populations, the recognition heuristic may underplay the possibility that for many numerical domains failure to recognize target items (such as countries) may be rare. The few empirical studies supporting one-reason decision models have tended to overlook how these models are likely to fit when recognition is uniformly high, or when multiple cues can be retrieved (for critical reviews see Broder & Schiffer, 2003; Hogarth & Karelaia, in press; Shanteau & Thomas, 2000).

If single predictive cues are less frequent than suggested by one-reason decision models, two further claims can be made that should impact how people make ordinal judgments. First, the possibility exists that people are sensitive to the internal validity of cues, as well as their ecological validity. Experiment 4 showed that people were able to identify and use predictive cues and reject poor ones, at least when the to-be-estimated items were representatively sampled. These data implied that cues are evaluated in terms of their ecological validity alone. Still, when items are unrepresentative, an evaluation of ecological validity is unlikely to result in the retrieval of cues capable of discriminating between target items, and is necessarily the case when target items belong to the same partition. This can be demonstrated by considering infant mortality rates for geographical regions rather than globally. Per capita GDP, for instance, only predicts infant mortality and life expectancy at $r_s < -.32$ for West European countries and is only marginally higher for Sub-Saharan countries where infant mortality rates are of real scientific and social interest. Asking this question decreases the accountable variance using wealth as the predictor from 76% to less than 15%. Furthermore, it is not obvious which single cue could predict infant mortality rates in these regions, since variation is likely due to any number of subtle differences between countries. It is imperative to consider that "unrepresentative" samples are not necessarily statistical anomalies that can be dismissed as improbable, but samples whose properties deviate significantly from normative sets, yet still have real-world relevance.

It was suggested that information might need to be evaluated for its predictive power in terms of the to-be-estimated items (i.e., internal validity) as well as ecological validity when conditions are less than optimal. Under what Shanteau and Thomas (2000) call *unfriendly* conditions, respondents are likely to assess the local validity of cues since additional information with a moderate to high internal validity could be useful in distinguishing between items, even if the information's ecological validity is not high. It is reasonable to suggest that people will recognize wealth as a poor cue when estimating infant mortality for just European countries while other cues might be considered relevant (e.g., type of health care systems).

The second claim is related to the first. Namely, that the cognitive architectures suggested by one-reason decision models appear too brittle to account for real-world performance (see Newell & Shanks, 2003, Newell, et al., 2003). As a result, these architectures tend to lack psychological validity, not least because of empirical evidence indicating that respondents often consider more than one item of information. Newell and Shanks (2003) demonstrated that the factors that influence the frequency of one-reason decisions include people's perception of cue validities, how well the cues are known, and how well structured the relationship is between cues and the target variable. For many domains cue validities are not well known, especially when items are unrepresentative. Broder (2000) has also pointed out that strong versions of one-reason decision-making are untenable in light of empirical evidence (see also Newell, et al., 2003). Experiment 2 supported the general conclusion that respondents often accessed more than one type of information, although wealth remained the dominant source of ordinal information. Ordinal-conversion describes a flexible, process oriented, framework that can account for both single cue and multiple cue use within the same overall architecture. This type of flexibility should be particularly relevant in unfriendly environments. The remainder of this section is given over to describing and comparing ordinal-conversion with more general decision models leading to predictions about performance and cue selection when the to-be-estimated items are unrepresentative.

Models that describe how information is integrated into decision processes are commonly classified as either compensatory or non-compensatory depending on the number of cues selected and how different sources of information interact. The compensatory/non-compensatory dichotomy assumes two elementary descriptions of how process and information combine (Broder & Schiffer, 2003; Hogarth & Karelaia, in press). The proxy/ordinal-conversion position cuts across this distinction. If the proxy cue is capable of predicting a partition and assigning a numerical value within that partition, then single proxy cues are the only information needed for a particular target item. When this occurs, cue selection is equivalent to the TTB heuristic. However, proxy cues may be used to predict a partition, but additional information may be required in order to decide on a specific numerical value. Figure 7a depicts how compensatory models, such as Equal Weighting or multiple regression models, assume that judgments are made on the basis of multiple cue combinations. Figure 7b on the other hand describes how proxy cues can be used to assign items to partitions, but also how other information may be used to assign specific values to discriminate items within a partition. It follows that when estimating infant mortality rates for West European countries, wealth provides strong reasons for assigning all target items to the low mortality partition, but other information is likely to be required because knowledge about wealth is subjective, often inexact, and tends to overlook factors that become more important when differences in wealth are relatively small.

Ordinal-conversion is non-compensatory because there is no impact between proxy cues and the values of any additional information used to generate specific numerical values. On the other hand, the probability that secondary information will be accessed is, theoretically dependent on how well the proxy information can discriminate between items. While this is not strictly compensatory, the overall impact of additional information is dependent on how proxy cues discriminate between items. Hence the utility of defining models as compensatory or non-compensatory appears less relevant than understanding how estimation strategy and cue use interact.

Still, there are no simple methods for determining individual participant's subjective understanding of wealth, a priori, without introducing experimental demands.

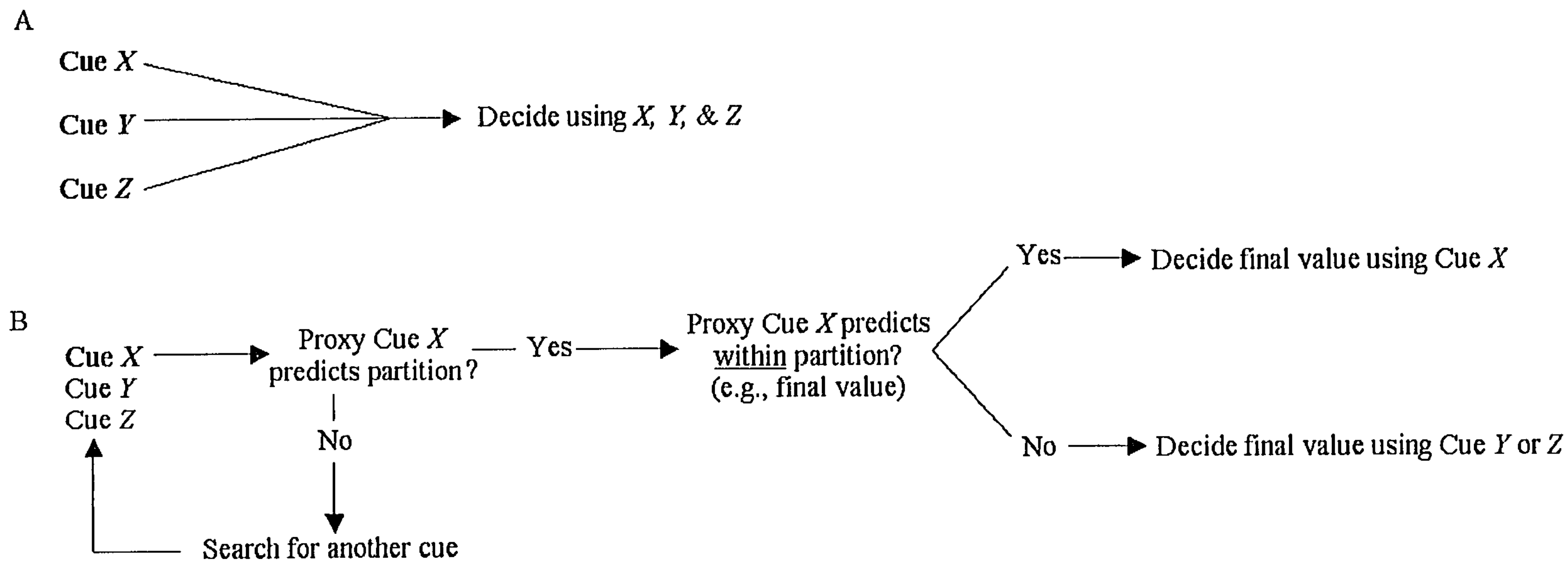


Figure 7. *Multiple cue and serial single-cue selection models.*

Moreover, the threshold of uncertainty associated with switching from single to two-cue solutions is unknown. As a result there are likely to be both between and within subject differences that reflect mixed reactions to different items, the types of information associated with them, as well as internal validity (Brown & Sinclair, 1999; Conrad, Brown, & Cashman, 1998; Lee & Brown, 2004, Siegler, 1987). For example, respondents may prefer to choose between two competing types of information, while others may decide that the additional cues can be used to discriminate between items in the same partition. These differences suggest that the impact of new information on rank order will consist of a weighted blend of different reactions. The aim of Experiment 5 was to examine whether people are sensitive to internal as well as ecological validity.

Two features have characterized the experiments presented so far. First, the criterion variable (infant mortality) is well predicted by a single ecologically valid proxy cue (wealth). Second, the target countries have been representatively sampled, so the internal validity of both predictive and non-predictive cues have matched their actual ecological validity. The following study uses the same procedure as Experiment 4. Here, respondents estimated infant mortality rates for a set of non-representative countries, where the internal validity between pcGDP and infant mortality for the to-be-estimated items was chosen to be extremely low, while population was now a moderate predictor of infant mortality. Of course, the ecological validity for both cues remains constant.

If people are responsive to internal cue validity then cue usage for population should increase significantly between Estimates 1 and 2 for participants in the population only group, a result that was not observed in Experiment 4. With wealth no longer a good predictor of infant mortality for the countries being estimated (despite the fact that wealth remains highly predictive for infant mortality across the globe), population should be judged by some participants as a viable alternative for assigning countries to partitions, or at least useful in discriminating between countries. For example, by inferring that higher populations are associated with higher population densities and increased infant mortality due to contagious diseases and unsanitary conditions. This effect should result in a significant increase in mapping accuracy due to the increased internal validity between population and actual infant mortality, despite the absence of any effect of population on mapping accuracy in Experiment 4.

Nonetheless, it is unlikely that internal validity will be a strongest or only determinant of cue usage. Pre-existing beliefs about the ecological validity of wealth are still likely to predict performance when accurate information is available. Consequently, the pcGDP only group should demonstrate a lower correlation between pcGDP and estimated mortality during Estimate 1 than found in Experiments 3 or 4, but show the same increase in cue use when actual wealth information is presented in Estimate 2. This pattern is predicted because proxy cues provide exact values (rather than estimated pcGDP) allowing a finer grained distinction between infant mortality rates during Estimate 2 (despite a predicted effect of decreasing mapping accuracy due to the low internal validity of wealth as a cue).

Alternatively, if people are unaware or unconcerned by internal validity then ecological validity will be the only factor that determines cue selection. Although accuracy should decrease, there are clear implications for the patterns of cue use. First, the availability of population information in the population only group should have no effect on mapping accuracy between the first and second estimates, since cues with low ecological validity should continue to be rejected. Second, this decrease in accuracy should be greater for the both-cues group, since these participants are expected to rely on the wealth information only. In summary, the absolute degree of cue usage may decrease compared with those reported in Experiment 4 (because respondents find it difficult to map estimates onto restricted cue ranges), but the overall pattern of performance should remain the same.

Experiment 5

Method

Participants

Seventy-two undergraduates (45 females and 27 males) participated for course credit. Median age was 19.

Materials

Thirty-four countries were chosen as targets. In this set actual pcGDP correlated with infant mortality at $r_s = -.03$, and actual population with infant mortality at $r_s = .43$. Hence, the ecological validity of wealth remained high but the internal validity was no longer predictive. Similarly, population was unrepresentative and had a moderate internal correlation with infant mortality. In order to control for covariance between the two cues, countries were selected so that actual pcGDP and actual population correlated at $r_s = .09$.

More generally, the countries were selected to be recognizable and global rather than from any specific region (Appendix IV). Although these items spanned the range of infant mortality rates (7-61/per 1000 births) and populations (0.8 million-1.2 billion), there was a constricted range of per capita incomes ($1,600 - $18,600). The aim was to select items that were likely to be less discriminable in terms of wealth, making it difficult to assign target items to partitions on the basis of wealth. This in turn should serve to increase the relevance of the internal validity given that there are few (if any) alternative cues with high ecological validity.

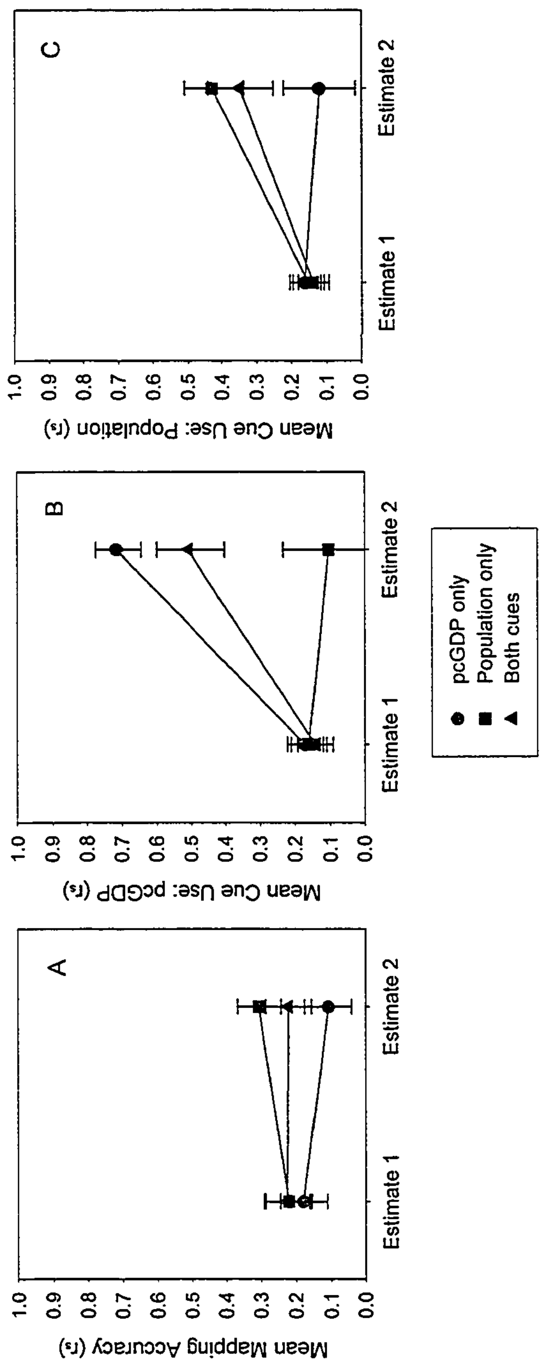
Results and Discussion

A preliminary analysis showed that there were no between groups differences for Estimate 1 in mapping accuracy, pcGDP cue use, or population cue use (Figure 8). All 2 (estimate) x 3 (group) mixed factor ANOVA returned significant main effects for estimate and group, as well as significant interactions for mapping accuracy, pcGDP cue use, and population cue use as dependent variables. A manipulation check showed that Estimate 1 accuracy was significantly lower than observed in Experiment 4, $r_s^{map1} = .21$, $t(142) = 17.73, p < .01$, indicating that participants found it difficult to discriminate between countries. As a result, the following analyses concentrate on interactions that differ from those found in Experiment 4 as well as specific within and between groups simple effects that highlight the theoretical points under consideration, especially those relating to the effects of internal and ecological validity.

pcGDP-Only

Experiment 4 showed participants tended to correlate estimates more highly with wealth than actual infant mortality during Estimate 1. It was predicted that participants in Experiment 5 would have difficulty discriminating between target countries on the basis of wealth and, as a consequence, find it difficult to assign infant mortality rates to individual target items using alternative cues. It follows that both mapping accuracy and the correlation between estimates and pcGDP was poor for the initial estimates. Nevertheless, cue usage was predicted to increase at Estimate 2 resulting in a

Figure 8. *Accuracy and cue use as a function of cue group.*



corresponding decrease in mapping accuracy due to a strong effect of ecological validity and the acceptance of wealth information in absence of alternative cues. A paired t-test showed a decrease in mapping accuracy significant at the .06 level, $t(23) = 2.00, p < .06$ (Figure 8a).

This was the only group to show any decrease in accuracy, and the effect can be attributed to an increase in pcGDP cue usage, $t(23) = -7.01, p < .01$. The direction and strength of this effect provides confirmatory evidence that ecological cue validity continued to have a large influence on cue selection (Figure 8b). It follows that ecological validity is likely to obscure any unique effect of internal validity. Even when the proxy cue has little internal validity, it seems that people's prior beliefs about the nature and causes of infant mortality continue to exert a strong influence on information selection. However, this does not preclude the notion that people may still be sensitive to internal validity. In order to test this assumption it is necessary to examine the population only and both cues conditions carefully since population had no impact on cue usage or accuracy when the internal validity of wealth was high in Experiment 4.

Population-Only

If ecological validity is the only factor determining cue selection then the population cues should have been ignored. Paired samples t-tests were conducted to test changes in accuracy and changes in population cue use (Figure 8a and 8c). The results confirmed a significant increase in accuracy, $t(23) = -2.07, p < .05$, that was also associated with an increase in correspondence between estimates and actual population, $t(23) = -3.77, p < .01$.

This finding is interesting as it provides tentative evidence that changes in both cue use and mapping accuracy are consistent with an effect of internal validity, and results that are inconsistent with Experiment 4 (where the population information was ignored and Estimate 2 correlated more highly with actual pcGDP than actual infant mortality). The proxy/ordinal-conversion position suggests that participants should treat the population information as potentially relevant under these conditions, by correctly concluding that the internal validity of wealth is low. Under experimental constraints that controlled for the availability and accuracy of cue information, internal validity was shown to impact how people treat cues with different ecological validities, and that ecological validity alone was not the only predictor of cue selection.

This is not to suggest that respondents are always aware that using population was likely to increase accuracy. Rather that subjective wealth can be used to assign the poor and developing countries to different partitions but may be poor at discriminating between countries belonging to the same partition. It is unlikely that participant's subjective understanding of wealth was fine grained enough to distinguish between many of the target countries. While proxy cues are useful because they are better known than the target domain, there are limitations to how well proxy cues discriminate closely related items, and whether differences between closely related target items can really be attributed to wealth alone. Alternatively, participants may have realized that relatively small differences in wealth are not likely to predict ordinal differences in infant mortality due to other factors. In any case, a more telling analysis of how respondents treated population information can be conducted by examining the both-cues group.

Both-Cues

In Experiment 4 the both cues group was found to map their responses using pcGDP information exclusively, resulting in performance that was almost indistinguishable from the pcGDP only group. Once again, a very different pattern of results was obtained in Experiment 5. Figure 8a reveals that there was no overall change in accuracy although cue use data indicated that performance was driven by both cues (Figure 8b and 8c), as opposed to pcGDP alone; in addition to a significant increase in pcGDP cue use, $t(23) = -5.70, p < .01$, there was also an increase in correlation between estimates and the population cues, $t(23) = -4.39, p < .01$.

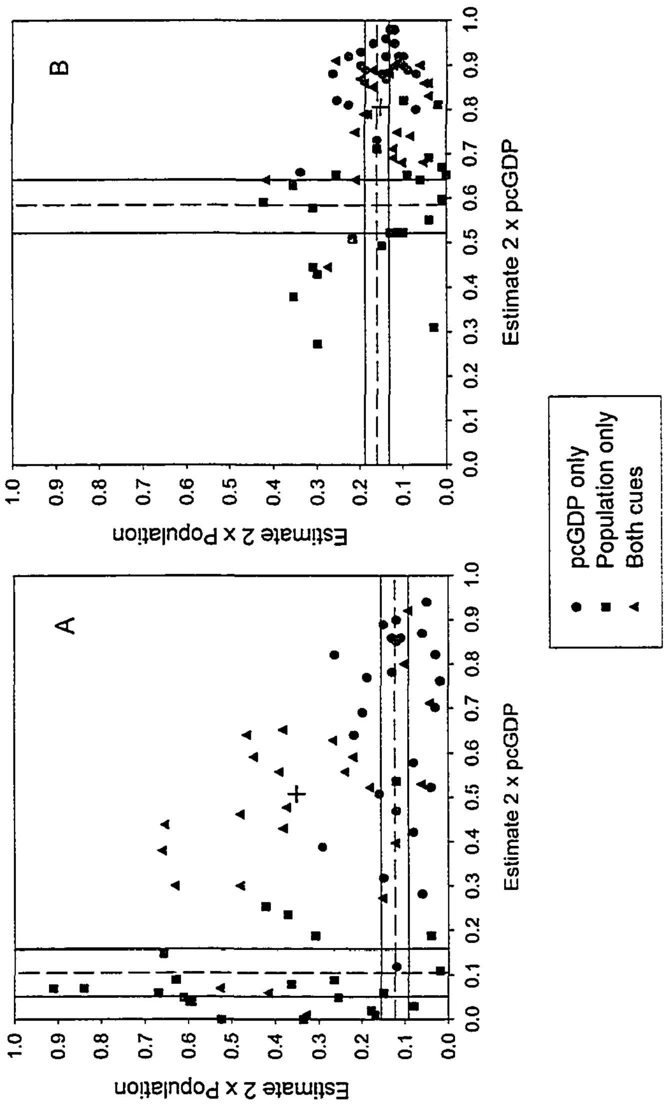
On the one hand the increase in pcGDP cue use can be explained by the strong tendency for ecological validity to influence cue selection. On the other hand, if ecological validity was the only criteria for selecting cues, then accuracy should have decreased and no effect of population should have been found (i.e., the same effects as obtained for the pcGDP only group, and a replication of the pattern found in Experiment 4). On the contrary, it appears that respondents did take into account population by employing both types of information; wealth, which had the effect of decreasing accuracy, and population, which increased accuracy since accuracy at Estimate 1 was poor to begin with. The net influence of using both cues would likely result in no change in accuracy, although the effect depends on the weighted average influence of each cue.

This interpretation is theoretically important by providing tentative evidence that people do engage in multiple cue selection, and that cue selection is mediated by internal validity. Holding the type of estimation task constant across Experiments 4 and 5 demonstrates that knowledge about infant mortality is not the only factor in cue selection. It seems that even cues with extremely poor ecological validity can be considered useful when ecologically valid cues do not provide adequate information to assign numerical values to closely related items from the same partition.

Still, a multiple strategies approach cautions against making a bold claim that people use multiple cues by averaging over data because the mean correlations overlook the possibility that some respondents used the pcGDP information while others used populations. On this view, a more detailed analysis can be conducted by comparing each participant's cue usage coefficients for pcGDP and population use. Figure 9a plots the 72 participant's cue use statistics for their second estimates in a two-dimensional space. Each participant's pcGDP x Estimate 2 correlation coefficient is plotted on the abscissa, and their population x Estimate 2 coefficient on the ordinal axis. The reference lines represent the mean correlation coefficients (and 95% confidence intervals) for the pcGDP and population only groups for the cues they *did not* receive (e.g., the vertical reference line indicates the mean correlation between pcGDP x Estimate 2 for the population only group).

The two-dimensional space permits a visual analysis of each participant's reaction to the presented information as a function of both cues, and a comparison of cue use with the other experimental groups. A pattern of results emerges that suggests three different reactions to the available information. First, there appears a group of participants (n = 7) whose estimates correlated poorly with the population cues (r_s < .2) and more highly with pcGDP, r_s^{gdp2} = -.78. Using r_s = .2 as the cut-off, a second group (n = 3) of participants appear to have relied on population, r_s^{pop2} = .42. Finally, a group of 14 participants whose estimates correlated higher than r_s = .2 on both cue types, r_s^{pop2} = .48 and r_s^{gdp2} = .56. Fisher's Exact Tests showed that the groups characterized by the r_s = .2 cut-offs were not significantly different from the pcGDP and population only cohorts. However, those

Figure 9. *Experiments 4 and 5: Cue use matrices as a function of group.*



defined as using both cues were significantly different from the two singleton groups on the cues these groups did not receive, $p < 0.05$, meaning these participants shared cue use characteristics of both singleton groups.

However, this interpretation is post-hoc and the data requires more principled comparisons before any substantive claims can be made.[6] A comparison between the both cues groups from Experiments 4 and 5 provides more robust evidence that most participants in Experiment 5 selected (and integrated) both types of information, and that this decision was mediated by internal validity. Figure 9b shows the cue use plots for the three cue groups in Experiment 4. These data reveal that these participants relied almost exclusively on pcGDP, although the population only group correlated their estimates less with pcGDP than those who received pcGDP information (this is simply the difference between subjective and accurate pcGDP information).

In summary, Experiment 5 demonstrated that internal validity has an effect on cue selection for cues with poor ecological validity when items are difficult to discriminate using an ecologically valid proxy cue. In this study the range of infant mortality rates and

[6] In a subsequent replication of the both cues group in Experiment 5, participant's provided post-estimate strategy reports describing how they selected cues during Estimate 2. These data confirmed the mixed reactions account: 40% reported using pcGDP and 52.5% both cues. More importantly, the pattern of cue use demonstrated the same characteristics found in Figure 9a, with three distinct groups whose actual cue use was consistent with their reported strategies, and therefore consistent with a mixed reaction explanation of single and multiple cue use under less-than-optimal conditions.

populations were consistent with a representative sample of countries, but the range of national per capita income was attenuated, resulting in a sample of countries where pcGDP did not predict infant mortality while population had moderate internal validity. It was hypothesized that difficulty in discriminating between items using the proxy cue might provide feedback alerting respondents to the possibility that wealth may have poor internal validity. For the pcGDP only group, the difficulty was compounded by the absence of alternative cues. In the absence of alternatives, ecological validity still predicted the selection of pcGDP when accurate information became freely available. The same cannot be argued for the population only group where population was, to varying degrees, accepted as potentially relevant and hence adopted as a source of information.

The both cues group faced a more interesting and difficult decision. Both pcGDP and population were openly available, only difficulty in assigning experiences during Estimate 1 should have indicated that wealth was not likely to prove the best predictor of infant mortality rates. Twenty-seven of the thirty-four target countries in Experiment 5 differed from their nearest fiscal neighbours by less than $1000 pcGDP. Wealth might be the most ecologically valid cue, but for many sets of countries there are many predictors that are better than national income (and able to off-set the influence of pcGDP). Still, not all participants interpreted this dilemma in the same way, and their reactions appear to reflect differing beliefs about the utility of using wealth and their decision to countenance population as viable cue, which in normal circumstances would be rejected. In contrast, the participants in Experiment 4 did not appear challenged with having to decide between selection strategies since their experience during Estimate 1 was not inconsistent with normative beliefs about differences infant mortality. The reasons behind these different selection strategies are likely to be profound because they reflect differences in the understanding and representation of real world beliefs. This raises the questions about individual differences, and possible characteristics of people who decide to choose one cue and those that choose multiple cues. The differences in selection strategies between Experiments 4 and 5 also suggest cognitive mechanism capable of monitoring performance and evaluating sets of items that provides feedback for cue selection but do not play a direct role in processes involved in numerical estimation.

Even so, this does not imply that some participants made better decisions in Experiment 5 than others, by choosing to use population for example. This type of qualitative interpretation overlooks the notion that adaptive behaviour is probably best served by holding previous beliefs in mind while examining possible alternatives. One alternative was to strike out using unproven population information. For most of the both cues group this concept of adaptability appeared to manifest itself as spreading the risk between different types of information. If our respondents choice to make use of the population information is a reflection of cognitive adaptability, then the results highlight an important alternative to the position taken by the ABC Research Group.

The final chapter will review the theoretical conclusions that can be drawn from the five experiments presented here, focusing on how this work contributes toward a better understanding of people's ordinal judgments, and the information used to generate them. In addition, implications for broader theoretical issue will be discussed such as cognitive architectures for dealing with more than one type of information and how cognitive adaptability may be better conceptualised as modification to existing strategies rather than positing the evolution of a "toolkit" of strategies.

Chapter VI

CONCLUSIONS

This project aimed to address a topic in real-world estimation concerning how people can rank order numerical estimates, even though the exact numerical values are often unknown. The main goal of this project was to describe, explain, and test the the ordinal-conversion strategy, as well as investigating the type and quantity of task relevant information people select to infer plausible answers. At a broader level, this project extends our understanding of the mapping component in the metrics and mapping framework that illustrates how people make use of metric and ordinal information to generate numerical estimates under uncertainty. Independence between metric and mapping accuracy has been demonstrated through dissociations showing that a subset of seed information can often improve people's metric accuracy (Brown & Siegler, 2001; Friedman & Brown, 1999; Lawson & Baghat, 2002; LaVoie, Bourne, & Healy, 2002) but rarely has any impact on mapping accuracy (c.f. Brown & Siegler, 1993).

Because numerical estimation often relies on plausible reasoning and the inferences derived from domain specific knowledge it is important to select a numerical domain that evokes the cognitive processes under investigation. Throughout this project participants were asked questions about infant mortality, on the basis that respondents were likely to have a good conceptual understanding about infant mortality and the factors that predict children's health, but unfamiliar with how mortality rates are typically measured or the metric properties of the domain. It was hypothesised that numerical domains exhibiting these characteristics should evoke selection of mapping information and processes described by ordinal-conversion, since a lack of metric knowledge leaves few alternative strategies other than rank ordering items by their relative values. Of course, there may be other types of numerical domain whose metric characteristics are better known, resulting in other estimation strategies described by Brown (2002) as numerical reconstruction or retrieval strategies. It was also suggested that in the absence of exact numerical information people would often access ordinally predictive cues. Experiments 1 and 2 showed that respondents' metric accuracy was indeed poor, but they were able to correctly identify wealth as the best predictor of infant mortality. Experiment 2 also revealed two findings that have important theoretical implications. First, that people's estimates correlated quite strongly with actual per capita Gross Domestic Product and that the strength of this association was higher than that between estimated and actual infant mortality, suggests that subjective wealth was better understood than actual mortality rates and likely to be the primary source of ordinal information. This is important because the data provided evidence for the ordinal conversion strategy and the use of proxy cues. Second, the verbal protocol analysis from Experiment 2 confirmed that wealth was the principle cue being used but many respondents also accessed additional information. The number of cues reported, and large individual differences, suggested that differences in knowledge about the numerical domain and specific target items is likely to account for differences in the number of cues accessed. Furthermore, these results provided grounds for arguing against the strong position proposed by the ABC research group regarding the prevalence of one reason decision making. This is not to suggest that the modal strategy is to use multiple cues. There are other factors, such as internal validity, that predict whether one or more cues will be selected. In consequence, cue selection in real world tasks is better described by a more moderate mixed reactions approach with emphasis on internal validity in addition to the structure of information in the environmental and cognitive considerations.

Initial evidence suggested that cue selection could be investigated and explained from a process focussed perspective, and thus contribute to the debate currently centred on one reason decision making, and in particular the QuickEst strategy. However, before a principled investigation could be undertaken a methodology was needed that was capable of isolating the effect of proxy cues. In reply, Experiment 3 introduced a technique called proxy cueing. Results showed that proxy cues influence ordinal accuracy, but not metric accuracy, provided that proxy information is available for all the target items. These data offered some initial evidence that proxy information provides item relevant rather than task relevant information due to the absence of any transfer effect on a subset of un-cued items. This was an important aspect of the project because the proxy cueing method provided a reliable technique for manipulating ordinal accuracy using ordinally relevant information, and made a principled investigation into the estimation processes associated with rank ordering items possible.

Experiments 4 and 5 compared the influence of ecological and internal validity on cue use and cue selection. Experiment 4 demonstrated that people are able to identify and use proxy cues effectively and are not influenced by irrelevant information, at least when the proxy cue is highly predictive. This result is encouraging for advocates of one reason decision making as it supports the premise that single cues can have large effects on ordinal accuracy and are likely to serve as the most important source of information, at least when the internal validity of the proxy cue matches the cue's ecological validity. Nevertheless, a very different pattern of results was obtained in Experiment 5 where multiple cue use (including the use of information with low ecological validity) was evoked when the internal validity of the proxy cue was poor. In conclusion, it appears that ecological validity predicts the selection of proxy cues, even when the cue lacks predictive power for the set of target items. The strong influence of ecological validity confirmed that, for most respondents, proxy cues were considered an important source of ordinal information, although a lack of internal validity predicts the use of additional information (if such information is available). Taken together, these findings make a number of significant contributions to the research area, and offer various directions for future research. In the remainder of thesis I will unpack each of these points in turn, while also examining issues that have not been addressed in this thesis and suggest new questions that have arisen out of the current findings.

Chapter I described an alternative position called QuickEst. It was made clear that the QuickEst heuristic was a radically different approach to numerical estimation than metrics and mapping, and ordinal-conversion in particular, despite the fact that both perspectives share some fundamental assumptions, such as the parsimonious use of task relevant information. QuickEst's primary assumption is that the assignment of items to partitions occurs through a process of elimination (of potential partitions) by aspects. Recall, that items are assigned to more fined grained partitions based on the predictability of a unique cue at each stage of elimination, and that items are finally assigned numerical values based on the mean (or prototypical) value for items belonging to the partition.

A further assumption is that QuickEst qualifies as a fast and frugal heuristic. Of course, fast and frugal is a relative concept. Nonetheless, the data from Experiment 2 indicates that the modal number of cues accessed is one, and that the median number of cues used is two. It is difficult to conclude that so few cues could provide enough discriminability to permit the assignment of items to sufficiently fine grained partitions as to result in rank order accuracy greater than .6. Based on the cue use data, it doesn't seem likely that the participants in the preceding experiments were able to retrieve enough task relevant information to make elimination by aspects a viable estimation strategy. Hausmann, Bröder, Pohl, and Läge (2005) have provided an empirical test of hypotheses predicted by the QuickEst heuristic. These researchers found no evidence for elimination by aspects, or secondary evidence that items with larger values on a given dimension are estimated more rapidly than items with small values (a prediction that falls naturally from the elimination by aspects hypotheses; QuickEst predicts that items are filtered into partitions moving from large values to small values, hence larger items should be estimated more quickly than smaller items).

A corollary to these arguments is that QuickEst seems neither as parsimonious in terms of the number of cues used, nor as quick as ordinal-conversion. Elimination by aspects demands a series of steps whose stopping rule is determined by an items membership to a partition based on cue membership. Ordinal-conversion on the other hand, cuts across the need to use a complete set of procedures for each target item in a set. Instead, participants are thought to set the metric and determine the range just once. The values generated in these initial procedures hold true across all of the to-be-estimated items (unless new metric information evokes a change in beliefs leading to a revisions), and need only be generated once for the first estimate. Moreover, the 45% of estimates in Experiment 2 that referred to one cue cannot easily be accounted for by QuickEst since the high rank order correlation between estimated and actual values could not be achieved using simple assignment in a two-partition solution. Indeed, the use of single cues as the modal response strategy indicates that the processes that underlie these estimates capitalize on some characteristic inherent to the proxy cue that provides enough discriminability so that a final numerical value can be chosen on the basis of the proxy cue alone. The proxy/ordinal-conversion position predicts that some estimates can and will be made on the basis of single proxy cues. For example, if it is believed that Somalia is the poorest nation on earth, then it becomes reasonably easy to assign the highest plausible infant mortality rate to Somalia, with no recourse to elimination by aspects. In summary QuickEst does not appear to be either as fast or as frugal an estimation strategy compared to proxy/ordinal-conversion. In addition, neither the empirical evidence reported by Hausmann, et. al. (2005), or the evidence presented here, supports elimination by aspects as psychologically plausible in numerical estimation.

Another important contribution has been to extend how we think about ordinal-conversion. Prior to this line of research, ordinal-conversion was thought to be a series of procedures that typically applied to most numerical estimates, as evoked by questions such as infant mortality rates. While the evidence presented here supports this nomothetic view, the overall picture has been refined as we observe individual differences in cue selection and how some responses appear to involve an extended set of cognitive processes used to include the use of more than one cue. The current argument in the judgment and decision making literature regarding the number cues people use to make plausible judgments has tended to focus on the psychological plausibility of Bayesian reasoning and multiple regression models using multiple cues versus arguments (largely promulgated by the ABC Research Group) advocating one-reason decision making. The evidence presented here tends to cut across this distinction. By taking a process approach this project has been able to uncover, to a limited extent, the degree to which individual respondents tend to use single or multiple cues, as well as within subject differences in cue use. This insight provides an interesting level of analysis that permits estimation researchers to move beyond the cue use argument. For example, the issue of individual differences raises the possibility that there exist systematic causes determining differences between respondents' cue selection strategies as well as unique causes that determine whether an individual target item will elicit one or more cues.

Placing theoretical comparisons aside, this projects main contribution is the discovery that internal validity plays an important role in cue selection. Conceptually, the proxy/ordinal-conversion position concurs with a Brunswikian perspective that ecological validity, and the structure of information in the environment, permits people to generate fairly accurate assessments about the characteristics of the target domain and to draw principled item specific inferences using task relevant information. However, changes in cue selection between Experiments 4 and 5 show that internal validity is also an important influence on the probability that proxy cues will be used alone, or possibly in conjunction with other cues. This is an interesting point worthy of further investigation as the effect of internal validity implies that respondents evaluate target sets as a whole, not just as single items. If this is the case, then numerical estimation may have two distinct cognitive components; those described in this thesis that deal with generating inferences and actual numerical estimates, and a second independent set of processes involved in monitoring performance and making evaluative judgments about the estimation task. The existence of a second set of processes dedicated to monitoring performance implies that evaluative information needs to be passed to estimation processes in order to influence cue selection.

Furthermore, this indicates that respondents are aware of certain aspects of their performance on numerical estimation tasks despite the fact that no feedback is provided. It seems that respondents may be able to infer something about their accuracy based on proxy cues alone (i.e., that poor performance is likely to ensue when sets of items seem to have similar values on the proxy cue variable). If future research shows that self-monitoring and task evaluation does occur, this will provide a new area of research for psychologists with interests not just in numerical estimation but across the decision sciences (assuming that monitoring and evaluation is a general component in judgment and decision making).

Finally, my investigation into ordinal-conversion is not yet complete, and there are still outstanding questions that could not be addressed within the scope of this thesis, especially those regarding cue selection when internal validity is poor. Experiment 5 indicated that respondents might be using both wealth and population information, at least for a significant proportion of their estimates. However, the evidence is inconclusive; the dual cue use explanation may be applicable to only a subset of respondents. Alternatively, respondents may be using wealth or population alone, but switching cues between trials (Footnote 6 briefly describes why these two explanations have been subsequently refuted). Additional research is also required to establish how well the proxy/ordinal-conversion approach describes numerical estimates in domains other than quality-of-life measures. For example, by extending the findings presented here in areas such as geography (i.e., latitude estimates) and pricing. Broadening our understanding of human estimation processes in real world tasks holds promising interdisciplinary opportunities, and it is my intention to continue working to expand our understanding in this area for the foreseeable future.

References

Allan, L. G. (1993). Human contingency Judgments: Rule based or associative? *Psychological Bulletin, 114*, 435-448.

Baird, J. C., Lewis, C., & Romer, D. (1970). Relative frequencies of numerical responses in ratio estimation. *Perception & Psychophysics, 8*, 358-362. Psychonomic Society, US.

Blair, E., & Burton, S. (1987). Cognitive processes used by survey respondents to answer behavioural frequency questions. *Journal of Consumer Research, 14*, 280–288.

Broder, A. (2000). Assessing the empirical validity of the "Take-the-best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 1332-1346.

Broder, A., & Schiffer, S. (2003). Take The Best versus simultaneous feature matching: Probabilistic inferences from memory and effects of representation format. *Journal of Experimental Psychology: General, 132*, 277-293.

Broniarczyk, S. M., & Alba, J. W. (1994). The importance of the brand in brand extension. *Journal of Marketing Research, 31*, 214–228.

Brown, N. R. (1995). Estimation strategies and the judgment of event frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1539-1553.

Brown, N. R. (2002). Real-world estimation: Estimation modes and seeding effects. In B. H. Ross (Ed.). *Psychology of Learning and Motivation: Vol. 41* (pp 321-360). New York: Academic Press.

Brown, N. R. & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review, 100*, 511-534.

Brown, N. R. & Siegler, R. S. (2001). Seeds aren't anchors. *Memory & Cognition, 29*, 405-412.

Brown, N. R. & Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. *Journal of Sex Research, 36*, 292-297.

Brunswik, E. (1952). The conceptual framework of psychology. In *International encyclopaedia of unified science, Vol. 1* (pp. 4-102). Chicago: University of Chicago Press.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally.

Cheng PW, & Novick LR. 1990. A probabilistic contrast model of causal induction. *Journal of Personality and Social Psychology, 58*, 545-567.

Centers for Disease Control. (2002, April 19). *Racial and Ethnic Disparities in Infant Mortality Rates: 60 Largest U.S. Cities, 1995-1998.* Retrieved April 29, 2005, from http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5115a4.htm

Central Intelligence Agency. (2005, April 29). *World Facts: Bangladesh.* Retrieved April 29, 2005, from http://www.cia.gov/cia/publications/factbook/geos/bg.html

Collins, A., & Michalski, R. (1989). The Logic of Plausible Reasoning: A core Theory. *Cognitive Science, 13*, 1-49.

Collins, A., Warnock, E.H., Aiello, N. & Miller, M.L. (1975). Reasoning from Incomplete Knowledge, in D. Bobrow & A. Collins (Eds.). *Representation and understanding*. New York: Academic Press.

Conrad, F., Brown, N. R., & Cashman, E. (1998). Strategies for estimating behavioral frequency in survey interviews. *Memory, 6*, 339-366.

Dawes, R. M. (1998). Behavioral decision making. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.). *The handbook of social psychology Vol. 1* (pp. 497-548). Boston: McGraw-Hill.

Dhami, M. K., & Ayton, P. (2001). Bailing and jailing the fast and frugal way. *Journal of Behavioral Decision Making, 14*, 141-168.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215-251.

Fasolo, B., McClelland, G. H., & Todd, P. (in press). Escaping the Tyranny of Choice: When fewer attributes make choice easier. *Marketing Theory.*

Friedman, A., &, Brown, N. (1999). Assesssing and Improving Geographical Beliefs: A Cognitive Approach. *Research in Geographic Education*, *1*, 1-13.

Friedman, A., & Brown, N. R. (2000a). Reasoning about geography. *Journal of Experimental Psychology: General. 129*, 193-219.

Friedman, A., & Brown, N. R. (2000b). Updating geographical knowledge: Principles of coherence and inertia. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 26*, 900-914.

Friedman, A., Brown, N. R., & McGaffey, A. (2002). A basis for bias in geographical judgments. *Psychonomic Bulletin and Review, 9*, 151-159.

Friedman, A., Kerkman, D., & Brown, N. R. (2002). Spatial location judgments: A cross-national comparison of estimation bias in subjective North American geography. *Psychonomic Bulletin and Review, 9*, 615-623.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review, 103*, 650-69.

Gigerenzer, G., Todd, P.M., & the ABC Research Group. (1999). *Simple Heuristics that Make Us Smart.* New York: Oxford University Press.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: strategies for qualitative research.* Chicago: Aldine.

Goldstein, D. G. & Gigerenzer, G. (1998). Recognition: How to exploit a lack of knowledge. In G. Gigerenzer & P. Todd (Eds.). *Simple heuristics that make us smart.* New York: Oxford University Press.

Hausmann, D., Bröder, A., Pohl, R. & Läge, D. (2005). QuickEst im Test: Eine empirische Überprüfung der Quick Estimation-Heuristik. AKZ-Forschungsbericht Nr. 9. Zürich: Angewandte Kognitionspsychologie.

Hertwig, R., Hoffrage, U., & Martignon, L. (1999). Quick estimation: Letting the environment do some of the work. In G. Gigerenzer & P. Todd (Eds.). *Simple heuristics that make us smart.* New York: Oxford University Press.

Hogarth, R. M., & Karelaia, N. (in press). Ignoring information in binary choice with continuous variables: When is less "more"? *Journal of Mathematical Psychology.*

Huttenlocher, J., & Hedges, L. V. (1994). Combining graded categories: Membership and typicality. *Psychological Review, 101*, 157-165. American Psychological Association, US

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Kahneman, D. & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80,* 237- 251.

LaVoie, N., Bourne, L. E., Jr., & Healy, A. F. (2002). Memory seeding: Representations underlying quantitative estimations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1137-1153.

Lawson, R. and Bhagat, P. S. (2002). The Role of Price Knowledge in Consumer Product Knowledge Structures. *Psychology & Marketing, 19*, 551-568.

Lee, P. J., & Brown, N. R. (2004). The role of guessing and boundaries on date estimation biases. *Psychonomic Bulletin & Review, 11*, 748-754.

Lindsay, R.M., & Ehrenberg, A.S.C. (1993). The design of replicated studies. *The American Statistician, 47*, 217-228.

Moshkovich, H. M., Mechitov,A., & Olson, D. L. (2002). Ordinal Judgments for Comparison of Multiattribute Alternatives. *European Journal of Operational Research, 137*, 625-641.

Murray, K. & Brown, N. R. (manuscript under review). Estimating and Updating Internal Reference Prices: An Examination of the Structure of Knowledge for Automobile Prices.

Newell, B.R., Weston, N.J., & Shanks, D.R. (2003). Empirical tests of a fast and frugal heuristic: not everyone “takes-the-best”. *Organizational Behavior and Human Decision Processes, 91*, 82-96.

Newell, B.R., & Shanks, D.R. (2003). Take the best or look at the rest? Factors influencing 'one-reason' decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 53-65.

Newman, J. W. (1977), Consumer external search: amounts and determinants. In A. G. Woodside, J. N. Sheth and P. D. Bennett (Eds.). *Consumer and industrial buying behavior*. Amsterdam: North Holland Publishing Company.

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776–783.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, UK: Cambridge University Press.

Paulos, J.A. (1990). *Innumeracy: Mathematical illiteracy and its consequences*. New York: Vintage.

Popper, K. R. (1934). *Logic of Scientific Discovery*. London: Hutchinson.

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review, 89*, 250-280.

Reder, L. M. (1987). Strategy-selection in question answering. *Cognitive Psychology, 19*, 90-134.

Rescorla RA, Wagner AR. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. W. Black and W. F. Prokasy (Eds). *Classical Conditioning II: Current Research and Theory*. New York: Appleton-Century-Crofts.

Shanks, D. R. (1995). *The psychology of associative learning.* Cambridge: Cambridge University Press.

Shanteau, J., & Thomas, R. P. (2000). Fast and frugal heuristics: What about unfriendly environments? *Behavioral and Brain Sciences, 23,* 762-763.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General 116,* 250-264.

Simon, H.A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69,* 99-118

Simon, H. A. (1956). Rational choice and the structure of environments. *Psychological Review, 63,* 129-138.

Simon, H. A. (1969) *The sciences of the artificial.* Boston: MIT Press.

Simon, H., et al. (1986). Decision making and problem solving. Research Briefings 1986: *Report of the Research Briefing Panel on Decision Making and Problem Solving,* National Academy of Sciences. Washington DC: National Academy Press.

Simon, H.A. (1979). *Models of thought. New Haven.* Connecticut: Yale University Press.

Svenson, O. (1996). Decision making and the search for fundamental psychological regularities: What can be learned from a process perspective? *Organizational Behavior and Human Decision Processes, 65,* 252-267.

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

Urbany, J. E., Dickson, P. R. & Wilkie, W. L. (1989). Buyer Uncertainty and Information Search. *Journal of Consumer Research, 16,* 208-215.

Wilson, T. D., Lisle, D. J., Schooler, J. W., & Hodges, S. D., Kristen J. K., & LaFleur, S. J. (1993). Introspecting about reasons can reduce post-choice satisfaction. *Personality & Social Psychology Bulletin, 19*, 331-339.

Wilson, T. D., & Schooler, J. W. (1991). Thinking too much: Introspection can reduce the quality of preferences and decisions. *Journal of Personality and Social Psychology 60*, 181-192.

Appendix I

Stimuli used in Experiment 1 including actual infant mortality rates and true pcGDP.

(n.b. figures have been rounded for clarity)

Country	Infant Mortality Rate	pcGDP (Can. $)
Japan	3	42,200
Sweden	3	38,300
New Zealand	6	30,300
Cuba	7	3,600
South Korea	7	27,900
Hungary	9	18,600
Poland	9	13,700
Russia	20	12,900
Colombia	22	9,800
Thailand	22	10,200
Bosnia	23	2,800
Algeria	38	8,700
Indonesia	38	4,700
Turkey	44	10,400
India	60	3,900
Kenya	63	1,600
Nigeria	71	1,300
Pakistan	77	3,300
Ethiopia	103	1,100
Afghanistan	142	1,200

Appendix II

Stimuli used in Experiment 3 showing actual infant mortality rates and true pcGDP.

(Denotes seed or subset cue item)*

Country	Infant Mortality Rate	pcGDP (Can. $)
Sweden	3	38,300
Japan	4	42,200
Netherlands	4	40,000
Norway	4	47,800
Switzerland	4	48,300
Singapore	4	38,300
Australia	5	37,200
Belgium	5	40,500
Canada	5	43,000
Germany	5	40,700
Ireland	6	42,400
Italy	6	37,700
New Zealand *	6	30,300
Cuba *	7	3,600
United States *	7	56,300
South Korea	8	27,900
Hungary	9	18,600
Poland *	9	13,700
Kuwait	11	23,400
Macedonia	13	6,800
Jamaica	14	5,700
Bulgaria	15	9,600
Uruguay	15	14,300
Dominican	17	9,000
Romania	19	10,600
Panama	20	9,200
Russia	20	12,900
Ukraine	21	6,500
Bosnia	24	2,800
Mexico	25	14,000
China	28	6,700

Appendix II

Country	Infant Mortality Rate	pcGDP (Can. $)
El Salvador	28	7,100
Iran	29	9,900
Thailand	30	10,200
Vietnam	30	3,300
Honduras	31	4,000
Nicaragua *	34	3,900
Brazil	37	11,500
Peru	39	7,400
Algeria	41	8,700
Turkey	47	10,400
Mongolia	53	2,700
Bolivia	59	4,000
Egypt	60	5,700
India	63	3,900
Zimbabwe	63	3,800
Cambodia	65	2,300
Kenya	68	1,600
Bangladesh	70	2,700
Nigeria	73	1,300
Nepal *	74	2,200
Pakistan	81	3,300
Uganda	91	1,900
Ivory Coast	94	2,400
Congo	100	900
Rwanda	119	1,600
Somalia	124	900
Afghanistan *	145	1,200

Appendix III

Stimuli used in Experiment 4 showing actual infant mortality rates, pcGDP, and population.

Country	Infant Mortality Rate	pcGDP (Can. $)	Population (Millions)
Sweden	3	38,300	8.9
Japan	4	42,200	126.8
Netherlands	4	40,000	16
Switzerland	4	48,300	7.3
Australia	5	37,200	19.4
Belgium	5	40,500	10.3
Canada	5	42,100	31.6
Germany	5	40,700	83
United Kingdom	6	38,300	59.6
Italy	6	37,700	57.7
New Zealand	6	30,300	3.9
Cuba	7	3,600	11.2
United States	7	56,300	278.1
South Korea	8	27,900	47.9
Hungary	9	18,600	10.1
Poland	9	13,700	38.6
Kuwait	11	23,400	2
Macedonia	13	6,800	2
Fiji	14	8,100	0.8
Jamaica	14	5,700	2.7
Bulgaria	15	9,600	7.7
Uruguay	15	14,300	3.4
Sri Lanka	16	5,000	19.4
Dominican Republic	17	9,000	8.6
Romania	19	10,600	22.4
Panama	20	9,200	2.8
Russia	20	12,900	145.5
Malaysia	20	14,000	22.2
Ukraine	21	6,500	48.8
Bosnia	24	2,800	3.9
Mexico	25	13,1000	101.9
Venezuela	25	9,500	23.9
El Salvador	28	7,100	6.2

Appendix III

Country	Infant Mortality Rate	pcGDP (Can. $)	Population (Millions)
Iran	29	9,900	66.1
Philippines	29	6,200	82.8
Thailand	30	10,200	61.8
Vietnam	30	3,300	79.9
Honduras	31	4,000	6.4
Nicaragua	34	3,900	4.9
Ecuador	34	4,700	13.2
Peru	39	7,400	27.5
Algeria	41	8,700	31.7
Turkey	47	10,400	66.5
Mongolia	53	2,700	2.7
Bolivia	59	4,000	8.3
Egypt	60	5,700	69.5
Zimbabwe	63	3,800	11.4
Cambodia	65	2,300	12.5
Kenya	68	1,600	30.8
Nepal	74	2,200	25.3
Pakistan	81	3,300	144.6
Madagascar	84	1,400	16
Uganda	91	1,900	24
Ivory Coast	94	2,400	16.4
Congo	100	9,900	53.6
Rwanda	119	1,600	7.3
Somalia	124	8,900	7.5
Afghanistan	145	1,200	26.8

Appendix IV

Stimuli used in Experiment 5 showing actual infant mortality rates, pcGDP, and population.

Country	Infant Mortality Rate	pcGDP (Can. $)	Population (Millions)
Cuba	7	3,600	11.2
Poland	9	13,700	38.6
Costa Rica	11	13,200	3.8
Jamaica	13	5,700	2.7
Fiji	13	8,100	0.8
Bulgaria	14	9,600	7.7
Lithuania	14	11,800	3.6
Argentina	16	18,600	37.4
Serbia	17	3,500	10.7
Romania	18	10,600	22.4
Malaysia	19	14,000	22.2
Russia	20	12,900	145.5
Ukraine	21	6,500	48.8
Thailand	22	10,200	61.8
Colombia	22	9,800	40.3
Bosnia	23	2,800	3.9
Mexico	24	14,000	102
Venezuela	24	9,500	23.9
Trinidad	25	14,000	1.2
Philippines	25	6,200	82.8
China	25	6,700	1273
N. Korea	26	1,600	22
El Salvador	27	7,100	6.2
Vietnam	31	3,300	79.9
Brazil	32	11,500	174.5
Ecuador	32	4,700	13.2
Dominican Republic	34	9,000	8.6
Peru	37	7,400	27.5
Algeria	38	8,700	31.7
Indonesia	38	4,700	228
Iran	44	9,900	66.1
Turkey	44	10,400	66.5
Saudi Arabia	48	16,400	22.8
South Africa	61	14,600	43.6