Journal of Experimental Psychology: General

Magnitude Comparison Extended: How Lack of Knowledge Informs Comparative Judgments Under Uncertainty

Oliver Schweickart and Norman R. Brown

Online First Publication, January 14, 2013. doi: 10.1037/a0031451

Journal of Experimental Psychology: General

0096-3445/13/$12.00 DOI: 10.1037/a0031451

Magnitude Comparison Extended: How Lack of Knowledge Informs Comparative Judgments Under Uncertainty

Oliver Schweickart and Norman R. Brown
University of Alberta

How do people compare quantitative attributes of real-world objects? (e.g., *Which country has the higher per capita GDP, Mauritania or Nepal?*). The research literature on this question is divided: Although researchers in the 1970s and 1980s assumed that a 2-stage magnitude comparison process underlies these types of judgments (Banks, 1977), more recent approaches emphasize the role of probabilistic cues and simple heuristics (Gigerenzer, Todd, & The ABC Research Group, 1999). In this article, we review the magnitude comparison literature and propose a framework for magnitude comparison under uncertainty (MaC). Predictions from this framework were tested in a choice context involving one recognized and one unrecognized object, and were contrasted with those based on the recognition heuristic (Goldstein & Gigerenzer, 2002). This was done in 2 paired-comparison studies. In both, participants were timed as they decided which of 2 countries had the higher per capita gross domestic product (GDP). Consistent with the MaC account, we found that response times (RTs) displayed a classic *symbolic distance effect*: RTs were inversely related to the difference between the subjective per capita GDPs of the compared countries. Furthermore, choice of the recognized country became more frequent as subjective difference increased. These results indicate that the magnitude comparison process extends to choice contexts that have previously been associated only with cue-based strategies. We end by discussing how several findings reported in the recent heuristics literature relate to the MaC framework.

Keywords: comparative judgment, fast-and-frugal heuristics, magnitude comparison, recognition heuristic, symbolic distance effect

Supplemental materials: http://dx.doi.org/10.1037/a0031451.supp

Comparison is a fundamental capacity of the human information-processing system. The application of comparison processes spans from low-level perceptual judgments—*Which card is brighter: or ?* (Cattell, 1902)—to relatively simple knowledge-based judgments—*Which number is larger: 4 or 9?* (Moyer & Landauer, 1967)—to high-level judgments and decisions involving similarity—*How similar are the United States and Canada?* (Medin, Goldstone, & Gentner, 1993), preference—*Should I get a BlackBerry or an iPhone?* (Lichtenstein & Slovic, 2006; Payne, Bettman, & Johnson, 1993), and inference—*Which city has the larger population: Munich or Leipzig?* (Gigerenzer & Goldstein, 1996). Even when people make judgments about the attributes of single target items, they frequently resort to comparative processing, evaluating a given target relative to a pertinent norm, standard, or context (Goffin & Olson, 2011; Kahneman & Miller, 1986; Mussweiler, 2003; Mussweiler & Epstude, 2009). In brief, there is widespread agreement that comparison plays a central role in the way people interact with their physical and social environments, and in the way they understand the world.

Although comparison is a basic cognitive *function*, it is not true that all comparisons are performed by the same basic cognitive *process*. Indeed, the diverse comparison literature (cited above and reviewed below) indicates that different tasks evoke different processes and that people sometimes use different strategies to perform the same task. Thus, a core set of issues in this area involves identifying the relevant comparison processes, specifying their conditions of use, and understanding their behavioral consequences (Beach & Mitchell, 1978; Marewski & Schooler, 2011; Payne et al., 1993; Rieskamp & Otto, 2006).

The present study was undertaken with these issues in mind, but it was also motivated by a marked and surprising disconnect in the comparison literature. Specifically, we observed that two generations of researchers have focused on the same comparison task, but have done so from different perspectives. The task itself is a simple one; participants are presented with pairs of items (e.g., Munich, Leipzig) and are required to compare them on a quantitative criterion dimension (e.g., population). In the 1970s and 1980s, psychologists studied this *paired-comparison task* from an information-processing perspective, focusing on the origins of several robust response time (RT) phenomena (e.g., the *symbolic distance effect*, see below) and their representational implications (Banks, 1977; Moyer & Dumais, 1978). Broadly speaking, this research converged on the notion that an iterative magnitude

Oliver Schweickart and Norman R. Brown, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada awarded to Norman R. Brown. We thank Peter Lee for his contributions to this project.

Correspondence concerning this article should be addressed to Oliver Schweickart, Department of Psychology, University of Alberta, P217 Biological Sciences Building, Edmonton, AB, Canada T6G 2E9. E-mail: oliver.schweickart@ualberta.ca

comparison process underlies these types of judgments. That is, magnitude values are retrieved (or generated) and compared until a response criterion is met.

More recently, researchers in the field of judgment and decision making have used a structurally identical task—the *two-alternative choice task*—to investigate one-reason decision making and a family of fast-and-frugal inference heuristics such as the *recognition heuristic* (Goldstein & Gigerenzer, 2002) and *take-the-best* (Gigerenzer & Goldstein, 1996). On this view, a comparative magnitude judgment *under uncertainty* is based entirely on values of single probabilistic cues (e.g., recognizing one pair member but not the other), and magnitude information, per se, is assumed to play no role in the judgment process.

It is true that these two research streams have had different emphases: The magnitude comparison literature has tended to focus on relatively well-known criterion dimensions (e.g., animal sizes, numbers), whereas the fast-and-frugal heuristics literature has concentrated on less well-known dimensions (e.g., city population). Still, it remains to be determined whether the nature of the underlying mechanism fundamentally changes with the nature of the comparison. More concretely, do people use a different set of processes (or different processing architectures) when they decide which of two animals is larger than when they decide which of two cities is more populous?

Here we report a normative study and two paired-comparison studies designed to answer this question. In both paired-comparison studies, people were presented with the names of two countries and were timed as they decided which one had the higher per capita gross domestic product (GDP). The critical comparisons involved pairs with one known country and one unknown country. For reasons developed below, the *magnitude comparison* (MaC) position and the *fast-and-frugal heuristics* (FFH) approach, specifically the recognition heuristic (Goldstein & Gigerenzer, 2002), make different predictions about the nature of RT and choice distributions for these types of comparisons. Before reporting these studies, we provide a brief overview of prior research on comparative magnitude judgments and position the two theoretical frameworks—FFH and MaC—within this field. Then, we describe how each approach deals with pairs involving one recognized and one unrecognized item and derive competing predictions for RT and choice.

Comparative Magnitude Judgments and the Distance Effect

Research on the processes and representations underlying comparative magnitude judgments has a long and rich history in psychology. One of the fundamental characteristics of these judgments, the *distance effect*, was first identified in psychophysical studies in which participants had to compare the magnitudes of two physical stimuli (e.g., Cattell, 1902; Henmon, 1906; Johnson, 1939). The distance effect refers to the relationship between the time required to make a comparative judgment and the distance between two stimuli on the judged dimension. That is, as distance increases, RT decreases. Furthermore, the error rate in comparative judgments typically increases as distance decreases.

Distance effects, as cognitive psychologists learned in the late 1960s, are by no means restricted to perceptual comparison (Moyer & Landauer, 1967). In what grew into an extensive research program on *symbolic comparison* (for overviews, see Banks, 1977; Leth-Steensen & Marley, 2000; Moyer & Dumais, 1978; Potts et al., 1978), numerous studies showed that paired-comparison tasks that required participants to retrieve magnitude information from memory yielded RT and error patterns similar to the ones observed in studies on perceptual comparison. Therefore, distance effects in memory-based comparative judgments are often referred to as *symbolic distance effects* (SDEs; Moyer & Bayer, 1976).

SDEs have been documented in a wide variety of knowledge domains, which include (a) comparisons on concrete dimensions such as the sizes of animals (e.g., Moyer, 1973) and other everyday objects (Paivio, 1975); (b) comparisons on abstract dimensions such as the magnitudes of numbers (e.g., Dehaene, Dupoux, & Mehler, 1990; Moyer & Landauer, 1967), time, quality, temperature terms (Holyoak & Walker, 1976), and the semantic goodness of words (A. Friedman, 1978); and (c) comparisons that involve some degree of uncertainty with respect to the position of target items along the criterion dimension, such as the military power of countries (Kerst & Howard, 1977) and the price of cars (Brown & Tan, 2011).

What lends further support to the notion that the distance effect is a prime marker of the comparison process is its ubiquity. The characteristic RT and/or error functions associated with the distance effect (a) are already present at very young ages (Duncan & McFarland, 1980; Huntley-Fenner & Cannon, 2000; Sekuler & Mierkiewicz, 1977; Siegler & Robinson, 1982; Temple & Posner, 1998; Xu & Spelke, 2000); (b) have also been documented in nonhuman species (Brannon & Terrace, 1998; Rilling & McDiarmid, 1965); and (c) are not limited to magnitude comparisons, per se, but also arise in comparative judgments related to, for example, relative location (Maki, 1981), order (Brown & Siegler, 1991; Marshuetz, 2005), and preference (e.g., Busemeyer & Townsend, 1993; Dashiell, 1937).

Over the last two decades, a new line of research on memory-based comparative judgments has emerged in the field of judgment and decision making. Here, researchers have studied comparative magnitude judgments *under uncertainty* to investigate the psychological reality of a set of simple inference heuristics (Gigerenzer & Goldstein, 1996; Goldstein & Gigerenzer, 2002; Schooler & Hertwig, 2005). These heuristics are part of a broader research program on fast-and-frugal decision making, which has, at its core, the notion of an "adaptive toolbox." According to this perspective, decision makers commonly select from a catalogue of content-sensitive, ecologically grounded heuristics that are characterized by a limited information search and an absence of information integration (Gigerenzer & Gaissmaier, 2011; Gigerenzer, Todd, & The ABC Research Group, 1999; Todd, Gigerenzer, & The ABC Research Group, 2012).

Despite the overlap between this new line of research and the symbolic comparison literature from the 1970s and 1980s, both proponents and critics of fast-and-frugal inference heuristics have neglected this earlier research and have come to conceptualize the judgment process (and the underlying processing architecture) in a fundamentally different way (Brown & Tan, 2011). We next describe the different process models for choice situations involving one recognized and one unrecognized object.

FFH Framework: The Recognition Heuristic (RH)

The RH states that if a person has to compare two objects with respect to a criterion, and only one of the two is recognized, then he or she should infer that the recognized object has the higher criterion value (Goldstein & Gigerenzer, 2002). Recognition is understood to serve as a probabilistic cue, and the heuristic exploits the fact that in many domains, recognition of the target objects (e.g., cities) is correlated with the criterion (e.g., population). The RH is a content-sensitive and domain-specific strategy in that it can only be applied if one object is recognized and the other is not, and it is only useful in environments in which recognition is predictive of the criterion.

The processing architecture of the RH (and other fast-and-frugal heuristics such as take-the-best) is provided by the *theory of probabilistic mental models* (Gigerenzer, Hoffrage, & Kleinbölting, 1991). According to this framework, a comparative magnitude judgment requires two serial stages. In an initial stage, an attempt is made to construct a *local mental model* (LMM) that consists of the direct retrieval or deduction of the correct answer to the comparison question.[1] This stage does not involve any inductive inferences and results in a 100% confident judgment. If an LMM cannot be activated, as is assumed to occur frequently in real-world knowledge domains, a *probabilistic mental model* (PMM) is constructed in a second stage.

A PMM involves the retrieval of a set of task-relevant probabilistic cues that are ordered according to their ecological validity. At this stage, people perform a serial search through this database. The search is terminated if a cue is found that has a positive value for one member of a pair, but not the other, and a choice is made solely on the basis of that particular probabilistic cue. That is, no further cues are interrogated, nor are any cues integrated. This strategy is referred to as take-the-best (Gigerenzer & Goldstein, 1996). In contexts with only one recognized pair member, the first cue that is processed is recognition—given that it is predictive—and because it discriminates between the two pair members, choice is based entirely on this single probabilistic cue (i.e., RH is a noncompensatory strategy). A critical point to note is that the binary cues in a PMM are entirely stripped of any magnitude information. Thus, magnitudes, per se, play no role at this stage of the judgment process.

More recently, proponents of the FFH approach have argued that the RH is applied as the default strategy in comparative judgments involving one recognized and one unrecognized object. However, they also claim that people "suspend" the RH under certain conditions (Gigerenzer & Brighton, 2009; Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2010; Pachur & Hertwig, 2006; Volz et al., 2006). According to this view, "the use of the recognition heuristic involves two processes: first, *recognition* in order to see whether the heuristic can be applied, and second, *evaluation* in order to judge whether it should be applied" (Gigerenzer & Brighton, 2009, p. 132). Several conditions—both on the item and domain level—have been proposed to trigger the suspension of the RH during the evaluation stage (for comprehensive overviews, see Pachur, Todd, Gigerenzer, Schooler, & Goldstein, 2011; Pohl, 2011). Among the suspension criteria are the following: (a) A person can retrieve "conclusive criterion knowledge" (Pachur & Hertwig, 2006, p. 997) that allows to answer the comparison question by deduction (i.e., an LMM is constructed; cf. Footnote 1); (b) the recognition cue in a given domain has a low validity (Pachur & Hertwig, 2006; Pohl, 2006); and (c) a low retrieval fluency of the recognized item in a given pair signals that the recognition cue might not be predictive in that particular instance (Marewski, Gaissmaier, et al., 2010; Marewski & Schooler, 2011). This last condition presupposes that the assessment of the speed with which the recognized pair member is processed—a potential index of the criterion (Hertwig, Herzog, Schooler, & Reimer, 2008; Schooler & Hertwig, 2005)—is used as a cue to evaluate the applicability of the RH.

Over the past decade, the RH has attracted significant attention and has generated much controversy (Hilbig, 2010; Marewski, Pohl, & Vitouch, 2010, 2011a, 2011b; Pachur, Bröder, & Marewski, 2008; Pachur et al., 2011). From a processing perspective, three issues have been particularly pertinent. First, the idea that any knowledge beyond recognition is ignored (at the PMM stage) has been very controversial from the outset, with one camp claiming to provide evidence in support of noncompensatory processing (e.g., Goldstein & Gigerenzer, 2002; Marewski, Gaissmaier, et al., 2010; Pachur & Hertwig, 2006; Volz et al., 2006) and an opposing camp claiming to provide evidence against it (e.g., Bröder & Eichler, 2006; Glöckner & Bröder, 2011; Hilbig, 2010; Hilbig, Erdfelder, & Pohl, 2010; Hilbig & Pohl, 2009; Hilbig, Pohl, & Bröder, 2009; Newell & Fernandez, 2006; Newell & Shanks, 2004; Oppenheimer, 2003; Pohl, 2006; Richter & Späth, 2006). Second, the RH is assumed to operate on the strictly binary output of a recognition judgment; however, the link between the recognition-based inference and the underlying recognition process has remained unclear (e.g., Dougherty, Franco-Watkins, & Thomas, 2008; Erdfelder, Küpper-Tetzel, & Mattern, 2011; Gigerenzer & Goldstein, 2011; Pleskac, 2007; Rosburg, Mecklinger, & Frings, 2011; Schooler & Hertwig, 2005). The third issue concerns the role of familiarity in the choice process. Whereas the RH is assumed to use the binary values of the recognition cue, other researchers have proposed that choice is based on the *comparison* of the relative familiarity of the two pair members (e.g., Dougherty et al., 2008; Honda, Abe, Matsuka, & Yamagishi, 2011; Newell & Fernandez, 2006; Schooler & Hertwig, 2005; but see Marewski, Gaissmaier, et al., 2010; Marewski & Schooler, 2011).

Despite the extensive research on the RH, the choice context in which the heuristic is applicable has, to our knowledge, never been related to or systematically investigated from a magnitude comparison perspective, an alternative framework of comparative magnitude judgments under uncertainty to which we turn next. Specifically, we first introduce the framework and then test its predictions against those of the RH. Finally, in the General Dis-

[1] Gigerenzer et al. (1991, p. 507) formulated three conditions for the successful construction of an LMM: "(a) precise figures can be retrieved from memory for both alternatives, (b) intervals that do not overlap can be retrieved, or (c) elementary logical operations, such as the method of exclusion, can compensate for missing knowledge." For pairs consisting of one recognized and one unrecognized object, only the third condition is relevant (Hilbig, Pohl, & Bröder, 2009), because the first two conditions require the retrieval of metric or ordinal information for *both* objects. For example, the third condition would apply when a person knows that an object is either the largest or the smallest of all objects in the comparison set. In this case, no inductive inferences are required to answer the question. This type of strategy is commonly referred to as an *end-anchor* or *end-term* strategy in the magnitude comparison literature (Banks, 1977).

cussion, we make an attempt to integrate the pertinent findings reported in previous studies on the RH with the MaC framework.

The MaC Framework

The MaC framework integrates process models of symbolic comparison (see below) with elements of *metrics-and-mappings*, a framework for understanding real-world quantitative estimation (Brown, 2002; Brown & Siegler, 1993). MaC rests on three main assumptions: (a) the central process that underlies comparative magnitude judgments *under uncertainty* is a magnitude comparison process similar to the one that underlies comparative magnitude judgments *under certainty*; (b) the comparison process operates on magnitudes that are generated by processes akin to the ones that have been identified in other magnitude judgment tasks, in particular real-world estimation; (c) comparative judgments involving only one recognized object are based on the same comparison process as those involving two recognized objects. These assumptions are explicated in turn.

The symbolic comparison literature converges on the notion that an iterative two-stage comparison process underlies memory-based comparative magnitude judgments (Leth-Steensen & Marley, 2000). However, the specific implementation of this comparison process varies somewhat between models because of differing assumptions about stimulus representation (analog versus propositional debate; cf. Banks, Fujii, & Kayra-Stuart, 1976; Dean, Dewhurst, Morris, & Whittaker, 2005; Holyoak, 1977; Kosslyn, Murphy, Bemesderfer, & Feinstein, 1977; Moyer, 1973; Moyer & Bayer, 1976; Paivio, 1975). According to propositional models, such as the *semantic coding model* (Banks, 1977; Banks et al., 1976), the first stage of the comparison process consists of the retrieval or generation of discrete semantic codes ("Small," "Medium," etc.) that specify the magnitude of each of the two stimuli on the criterion dimension. Then, in a second stage, these codes are compared. If they differ, a response is initiated; if they match, further processing is undertaken to retrieve or generate more detailed codes. This retrieval-comparison cycle continues until the codes discriminate between the two stimuli (or the cycle is terminated).

Analog models, however, typically involve the computation of a magnitude difference, and the comparison process is often formalized by means of sequential sampling models[2] (e.g., Birnbaum & Jou, 1990; Buckley & Gillman, 1974; Link, 1990; Link & Heath, 1975; Moyer & Dumais, 1978; Poltrock, 1989). For example, according to Moyer and Dumais' *scan-plus-comparison model*—a member of the *random walk* class—analog stimulus values for both pair members are retrieved from memory first, and then the magnitude difference between the two is computed and added to a counter. Depending on the direction of the difference (i.e., positive or negative), the counter increases or decreases in value. This sequential process continues until the counter exceeds a predefined upper (positive) or lower (negative) boundary that triggers the execution of the corresponding response.

Both implementations of the magnitude comparison process account for the SDE with the same basic rationale. According to the semantic coding model, SDEs emerge because two stimuli that are far apart on the judged dimension require fewer retrieval-comparison cycles to generate nonidentical semantic codes than two stimuli that are close together. In the scan-plus-comparison model, however, SDEs arise because more iterations are required to reach one of the response thresholds as the difference between stimulus values decreases.

How are magnitude values generated in comparisons under uncertainty? We suggest that the processes underlying magnitude generation are equivalent to the ones that have been identified as operating in quantitative estimation (Brown, 2002; Brown & Siegler, 1993). In estimation tasks (e.g., What is the current per capita GDP of Nepal?), people often rely on an *ordinal-conversion* strategy when they are unable to retrieve the numerical value of a target. That is, they first determine the relative magnitude of the target item (e.g., "Nepal has a low per capita GDP") and then map this ordinal value onto the numerical response range to select the appropriate value (e.g., "C$1000"). It is likely that the processes that underlie the initial ordinal component of the estimation process also play an important role in the first stage of the magnitude comparison process. The estimation literature indicates that people typically rely on a weighted blend of two sources of information in the magnitude-generation process—domain-specific knowledge and familiarity-based intuitions (Brown, 2002; Brown & Siegler, 1993). Specifically, the following list identifies some of the most relevant processes. (a) *Ordinal retrieval*—A fact is retrieved that specifies the ordinal value of the target item. For example, when making an ordinal judgment about the per capita GDP of, say, Switzerland, a person might simply be able to retrieve the fact that Switzerland has a relatively high per capita GDP. (b) *Ordinal reconstruction*—Task-relevant information relating target item(s) and target dimension is retrieved in the course of a *plausible reasoning* process (Collins & Michalski, 1989). For example, a person might recall that Switzerland hosts several large banks and chocolate and watch manufacturers, which suggests that it has a relatively strong economy and probably a high per capita GDP. (c) *Categorical knowledge and categorical inference*—Complex real-world knowledge is often categorical in nature, and there is ample evidence that categorical information plays an important role in both estimation and paired-comparison tasks (e.g., Brown & Siegler, 1991; A. Friedman & Brown, 2000a, 2000b; Maki, 1981, 1982; Sailor & Shoben, 1996). Categorical knowledge seems to be particularly relevant in judgments about attributes of countries (W. J. Friedman & deWinstanley, 2006). For example, a person's knowledge about the distribution of economic well-being around the world might be organized by geographic regions, reflecting, for instance, that most Western European countries are relatively prosperous and most African countries are relatively poor. Category membership can thus serve as a determinant of a country's ordinal per capita GDP value. (d) *Memory assessment*—In addition to task-relevant domain-specific knowledge, familiarity-based intuitions have also been demonstrated to play an important role in the ordinal judgment process, in particular when domain-specific knowledge is sparse (Brown, Cui, & Gordon, 2002; Brown & Siegler, 1992, 1993). That is, people sometimes provide larger estimates for familiar items than for less familiar items even

[2] See Pleskac and Busemeyer (2010) for a more recent sequential sampling model in which a common underlying process is used to account for the interrelationship of confidence, RT, and choice in both perceptual and general knowledge comparison tasks. For approaches that integrate sequential sampling models with the PMM framework, see M. D. Lee and Cummins (2004) and Newell and Lee (2011).

though their actual values are the same; this suggests that implicit factors also influence ordinal judgments.

Finally, and crucially for the present study, research on real-world estimation has shown that estimates for unknown target items are often highly systematic (Brown, 2002; P. J. Lee & Brown, 2004). For example, when people are asked to estimate the populations of countries that are unknown to them, they tend to infer that these countries have relatively small populations. A similar type of inference likely occurs in paired comparisons involving one unrecognized object. Therefore, we expect that participants in the present set of studies reason that unrecognized countries have relatively low per capita GDPs. We speculate that these types of inferences are based on the same sources of information as the ordinal judgments for recognized items. First, unrecognized items form the "tail end" of the familiarity variable (e.g., Brown, Rips, & Shevell, 1985; Lichtenstein, Slovic, Fischhoff, Layman, & Combs, 1978). Thus, if familiarity-based intuitions are used as an index of the criterion, unrecognized items should be judged to have low criterion values (if familiarity and the criterion are positively correlated). Another possibility is that a more explicit meta-cognitive process is applied in which the *lack of knowledge* is interpreted as being systematically related to the criterion dimension (Collins, Warnock, Aiello, & Miller, 1975; Gentner & Collins, 1981). For example, a person might reason: "I know most of the wealthy countries in the world. The fact that I have never heard of Burkina Faso suggests that it is not a wealthy country, because I'd probably know about it if it were." Note that both mechanisms converge on the same ordinal judgment, namely that unrecognized items have relatively small criterion values (in domains in which recognition and the criterion are positively correlated).

Competing Predictions for RTs and Choice Patterns

The MaC and RH accounts make different predictions with respect to both RT and choice patterns in comparisons involving one recognized and one unrecognized item (henceforth *RU pair*). In terms of RT, the critical prediction of the MaC account is the characteristic SDE function. Specifically, in the per capita GDP domain, choice times should decrease as the subjective difference in per capita GDP between the recognized and unrecognized country increases.[3] This is because the unrecognized country in an RU pair is typically inferred to have a relatively low per capita GDP, and the magnitude comparison process will require more iterations to reach a response criterion when the unrecognized country is paired with a recognized country that also has a subjectively low per capita GDP (e.g., Burkina Faso vs. Cambodia) than when it is paired with a recognized country that has a subjectively medium or high per capita GDP (e.g., Burkina Faso vs. Ukraine).

In contrast, the RH predicts no relationship between choice time and subjective distance once systematic variability due to recognition time differences is taken into account. This prediction is motivated as follows: In general, there is little reason to assume that the evaluation stage of the binary recognition cue is a source of systematic RT differences between RU pairs (Hilbig & Pohl, 2009). However, this is not true for the preceding recognition stage. Given that retrieval time can be correlated with the criterion (e.g., Hertwig et al., 2008), items with higher criterion values might be retrieved faster, which, in turn, would also affect the overall RT for the comparative judgment (Gigerenzer & Goldstein, 2011). This prediction is also consistent with recent ACT-R implementations of the RH (Marewski & Mehlhorn, 2011; Marewski & Schooler, 2011), which show that the recognition time of the recognized pair member and overall inference time are commonly positively correlated. However, given that the RH is a noncompensatory strategy, once recognition time differences are taken into account, there is no clear reason why comparison RTs should still vary as a function of subjective distance.

The two accounts also make different predictions with respect to the frequency with which the recognized item in an RU pair is chosen (a response we refer to as *R-choice*). According to the MaC account, the proportion of R-choices should vary as a function of subjective magnitude difference. For comparisons in the per capita GDP domain, this means that R-choices should become more common as the subjective per capita GDP of the recognized country increases, given that the unrecognized country is consistently inferred to have a relatively low criterion value. The likelihood of choosing the recognized or unrecognized country should be about the same in cases in which both countries in the RU pair are inferred to have low per capita GDPs (assuming no response bias). Finally, when the subjective per capita GDP of the recognized country is lower than the subjective value of the unrecognized country, the unrecognized country should be chosen more frequently (a response we refer to as *U-choice*).

In contrast, the "classic" RH account (Goldstein & Gigerenzer, 2002) predicts no relationship between the proportion of R-choices and subjective magnitude difference. That is, the recognized country in an RU pair should be chosen consistently at a high rate. However, more recently, it has been proposed that retrieval fluency can serve as an RH suspension criterion (Marewski, Gaissmaier, et al., 2010). That is, when a recognized pair member is retrieved slowly, it can cause a suspension of the RH. Thus, the state-of-the-art RH predicts no relationship between choice and subjective distance, once the effect of retrieval fluency is taken into account.

Study 1

In this initial study, we examined the knowledge domain of national per capita GDP in some detail. We asked participants to rate their knowledge and estimate the current per capita GDP of various countries from around the world. The rationale for this was threefold. First, we wanted to determine how many of the world's countries Canadian university students typically recognize and whether recognition can serve as a useful predictor of per capita GDP—the latter being critical for the applicability of the RH. Second, the study enabled us to assess the overall accuracy of people's beliefs about the per capita GDP of nations and also provided us with valuable stimulus norms. Furthermore, it allowed us to determine what sources of information—familiarity-based intuitions and domain-specific knowledge—people typically rely on in making judgments about the magnitude of a country's per capita GDP. Third, we also wanted to gain a better understanding

[3] Obviously, the MaC account also predicts distance effects for pairs consisting of two recognized items (cf. Brown & Tan, 2011), but this prediction is not tested in this article.

of people's inferences about the relative per capita GDP values of unknown countries—a crucial issue for MaC.

The study consisted of two tasks. In the knowledge rating task, participants were asked to rate their knowledge of various countries on a 10-point scale. A rating of 0 indicated that a country was not recognized, and ratings greater than 0 indicated that a country was recognized, with higher values reflecting greater amounts of knowledge (i.e., 1 = *I recognize the name of the country but I know nothing about it* and 9 = *I know a great deal about the country*). The use of this scale allowed us to categorize each country as being either recognized or unrecognized, and simultaneously provided us with a measure of familiarity. This is because the familiarity (or fluency) of a target item is, in part, a reflection of its frequency of occurrence in the environment (Schooler & Hertwig, 2005). As continuous knowledge ratings are often highly correlated with environmental occurrence frequencies, they are commonly used as a measure of familiarity (e.g., Brown & Siegler, 1992, 1993; Honda et al., 2011). In the estimation task, the names of the countries were presented for a second time, and participants had to estimate the current per capita GDP of each to the nearest 100 Canadian dollars.

Method

Participants. One hundred twenty-two introductory psychology students participated in the study in exchange for partial course credit. All participants in this and the following studies were born in Canada and had English as their native language.

Materials. A total of 154 countries served as stimuli in the study. This set included virtually[4] all of the world's countries with a population greater than 750,000 at the time of testing (Central Intelligence Agency, 2008a, 2008b). From this pool of countries, a stimulus set consisting of 77 countries was created for each participant; the 154 countries were rank ordered according to their actual per capita GDP, and one country was randomly selected from each successive pair of rank-ordered countries. This sampling procedure ensured that each stimulus set included countries across the entire range of possible per capita GDP values.

Procedure. Participants were tested individually, and all responses were collected by a computer. During the knowledge rating task, country names were displayed individually in the center of the screen with an input field below. Participants were instructed to rate their knowledge of each country on the 10-point scale described above, to enter their response on the keyboard, and to confirm it by pressing the *enter* key.

At the beginning of the estimation task, all participants were informed of the current per capita GDP of the United States (C$58,000 at the time). This was done to reduce extreme metric inaccuracy. Furthermore, the instructions stressed that participants were not expected to know the exact answers but that they should "try their best" to estimate the per capita GDP of each country as accurately as possible. During the task itself, the same countries were presented as in the knowledge rating task. Participants initiated each trial by pressing the *backspace* key; this caused a country name to appear in the center of the screen. Once they had decided on a per capita GDP estimate for a presented country, they pressed the spacebar, which allowed them to enter the numerical response into an input field below the country name. After entering their response, participants pressed the *enter* key. The presentation order of countries was randomized separately for each participant and each task.

Results and Discussion

Because the actual per capita GDP of the United States was presented as a reference value in the instructions, data for this country were excluded from all analyses. The recognition rates, mean knowledge ratings, and median per capita GDP estimates for the remaining 153 countries are listed in Table S1 in the Supplemental Materials.

Recognition rates. On the basis of the dichotomized knowledge ratings, we computed recognition rates for both subjects and items. In general, participants recognized a median of 85% of the countries in their respective stimulus sets (Interquartile range [*IQR*] = 12%). This value is roughly consistent with the ones obtained in similar studies conducted with U.S. college students in the past (W. J. Friedman & deWinstanley, 2006; Fryman & Wallace, 1985). On an item level, recognition rates were more variable and also revealed a high degree of skewness (*Mdn* = 98%, *IQR* = 24%). That is, 69 of the 153 countries had perfect recognition rates of 100%, whereas the recognition rates for the remaining 84 countries ranged from 99% down to 13%.

Estimates. The accuracy of the per capita GDP estimates was analyzed on two different levels: metric accuracy and mapping accuracy (Brown, 2002; Brown & Siegler, 1993). *Metric accuracy* reflects the degree to which participants' beliefs about the statistical properties of the target dimension are accurate. It is commonly measured by the *signed order of magnitude error* (*SOME*) and the *order of magnitude error* (*OME*) (Brown & Siegler, 1992; Nickerson, 1981). The *SOME* statistic is defined as: $SOME = \log_{10}(\text{Estimated Value}/\text{Actual Value})$. It describes the discrepancy between the estimated and actual value of a target item in terms of orders of magnitude. An *SOME* value of 0 indicates that the estimate was perfectly accurate, a positive *SOME* value indicates that the estimate was too high, and a negative *SOME* value indicates that the estimate was too low (e.g., an *SOME* value of 1 means that the estimate was one order of magnitude too high). *OME* is simply the absolute value of *SOME* ($OME = |SOME|$). A participant's average *SOME* reflects the overall metric bias, whereas the average *OME* indicates the absolute metric error. *Mapping accuracy*, however, reflects the degree to which the relative magnitudes of items along the target dimension are accurate. It is typically measured by Spearman rank-order correlations between estimated and actual values.

For each participant, the median *SOME* and *OME*, as well as the rank-order correlation between estimated and actual per capita GDPs, were computed. The median *OME*s indicate that metric accuracy was, overall, rather poor (*Mdn* = 0.42, *IQR* = 0.17), and the median *SOME*s suggest that participants typically overestimated the per capita GDPs of countries by about one third of an order of magnitude (*Mdn* = 0.30, *IQR* = 0.35). The by-participant

[4] A small number of items listed as countries in the *World Factbook* were excluded or modified: (a) *Gaza Strip*, *Hong Kong*, *United Arab Emirates*, and *West Bank* were excluded; (b) *Democratic Republic of the Congo* was shortened to *Congo*, and the *Republic of the Congo* was excluded; (c) *Burma* was changed to *Myanmar*, and *Côte d'Ivoire* to *Ivory Coast*.

rank-order correlations revealed that mapping accuracy was moderately good (*Mdn* r_S = .55, *IQR* r_S = .16). Taken together, these results suggest that knowledge of national per capita GDPs in our participant population was generally rather limited in an absolute (metric) sense, but somewhat better in a relative (mapping) sense. Even on that level, however, it was far from perfect. Thus, the present knowledge domain should be well suited to study inductive reasoning processes.

Next, we assessed the relative contribution of familiarity-based intuitions and domain-specific knowledge in the estimation of national per capita GDP. To determine this, we tested whether knowledge ratings (our index of familiarity) and actual per capita GDP values (our index of domain-specific knowledge) were, independent of one another, correlated with estimated per capita GDP values. By-participant partial rank-order correlations between estimated values and rated knowledge, controlling for actual per capita GDP, indicate that participants relied in part on familiarity-based intuitions in the estimation process (*Mdn partial* r_S = .34, *IQR partial* r_S = .24, 95% bootstrap CI for the median [0.28, 0.38]). That is, even after controlling for actual per capita GDP, participants tended to provide higher estimates for well-known countries than for little-known countries. Conversely, partial rank-order correlations between estimated and actual per capita GDPs, controlling for rated knowledge, suggest that domain-specific knowledge also played a major role in the estimation process (*Mdn partial* r_S = .47, *IQR partial* r_S = .18, 95% bootstrap CI for the median [0.41, 0.49]). These findings are consistent with the results of previous studies in which participants had to estimate quantities associated with countries (e.g., population, land area), which also indicate that some blend of familiarity-based intuitions and domain-specific knowledge underlies the ordinal component of the estimation process (Brown, 2002; Brown et al., 2002; Brown & Siegler, 1992, 1993).

Ecological analysis. In order to determine whether recognition is a potentially useful cue in the per capita GDP domain, we examined, in a manner analogous to Goldstein and Gigerenzer's (2002) analysis, the relationship between countries' recognition rates, environmental frequencies (as measured by *Canadian Newsstand* (2009)[5] citations counts), and actual per capita GDP values. In addition to these three variables, we also included the mean knowledge ratings—our measure of familiarity—and the median per capita GDP estimates of each country in this analysis (see Figure 1). For the sake of consistency, we report Spearman rank-order correlations for all relevant pairs of variables.[6]

According to Goldstein and Gigerenzer (2002), environmental frequency functions as a mediator between recognition and the criterion. As shown in Figure 1, the *ecological correlation*, that is, the correlation between countries' environmental occurrence frequencies and their actual per capita GDP values, was .51, and the correlation between occurrence frequencies and recognition rates, the *surrogate correlation*, was .84. The usefulness of the RH depends on the strength between recognition and the criterion. The correlation between recognition rates and actual per capita GDPs, the *recognition correlation*, was .44. This substantial correlation indicates that recognition can serve as a useful predictor of per capita GDP. Thus, the domain should be well suited for the application of the RH. It is worth noting, however, that familiarity (as measured by rated knowledge)—one of the two critical sources of information according to MaC—is also a potentially useful predictor of the criterion (r_S = .49).

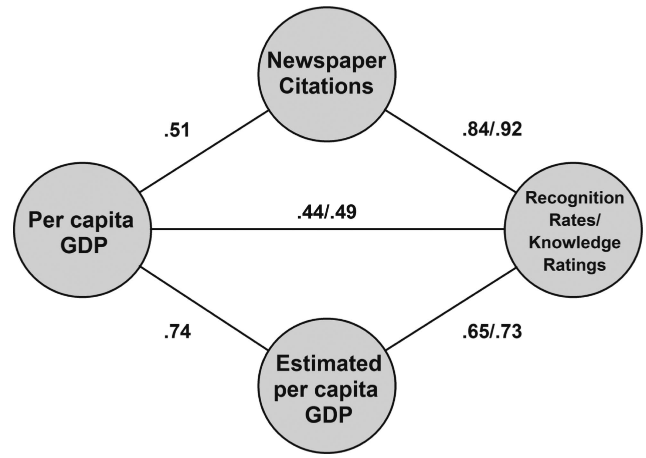

Figure 1. Ecological analysis of recognition and familiarity in the per capita gross domestic product (GDP) domain. The diamond depicts the interrelationship between the citation frequencies in *Canadian Newsstand*, actual per capita GDPs, recognition rates, mean knowledge ratings, and median estimated per capita GDPs for 153 countries. Numerical values represent Spearman rank-order correlations.

Subjective per capita GDPs of unknown countries. The left panel of Figure 2 shows the frequency distribution of per capita GDP estimates separately for recognized and unrecognized countries. The distribution for the unrecognized countries has, as expected, a reverse J-shape, indicating that unrecognized countries were typically estimated to have relatively low per capita GDPs. In fact, almost three quarters of the estimates for unknown countries fall at or below a value of C$20,000. Furthermore, the frequency distribution for the recognized countries also has a reverse J-shape, though a less steep one, which indicates that participants were generally aware of the skewed global distribution of economic well-being. Because relative, rather than absolute, magnitudes are of primary interest in the present set of studies, it is more informative to examine the frequency distribution of participants' rank-ordered per capita GDP estimates. This is because in the frequency distribution of ranked estimates, any variability due to between-subject differences in metric beliefs is eliminated. The right panel of Figure 2 shows that within the range of the 77 possible ranks, the ranked estimates for unrecognized countries fall below the midpoint of the range far more often than they fall above it. Furthermore, the shape of the distribution indicates that participants most frequently inferred that the per capita GDPs of unknown countries have low-to-medium magnitude values.

To summarize, the results of the present study indicate that the domain of national per capita GDP should be well suited for the

[5] *Canadian Newsstand* is a corpus of major Canadian newspapers. For each country, the number of citations was counted between January 1, 2008 and January 1, 2009. Citation counts for homonyms (e.g., China, Turkey, Jordan) and for countries with multiple common names (e.g., United Kingdom) were adjusted.

[6] Note that the relationship between some of these variables was analyzed on a subject level in the previous subsection. The corresponding item-level correlations are inflated because they are based on averaged values.

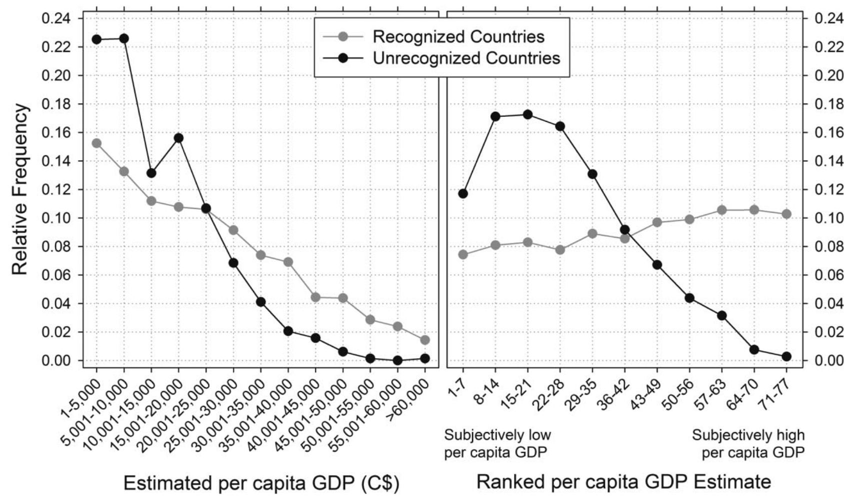

Figure 2. Grouped frequency distributions of per capita gross domestic product (GDP) estimates (left panel) and ranked per capita GDP estimates (right panel) for recognized and unrecognized countries. C$ = Canadian dollars.

study of inductive reasoning processes because participants' knowledge of this criterion dimension is rather limited. Furthermore, the per capita GDP domain fulfills the requirements for the application of the RH. Specifically, we found that participants typically recognize only a subset of the world's countries, and an ecological analysis revealed that recognition can serve as a useful predictor of the criterion. Finally, we obtained evidence that participants rely on familiarity-based intuitions and domain-specific knowledge when determining the relative magnitude of a country's per capita GDP and that unrecognized countries are typically inferred to have values in the low-to-medium magnitude range.

Study 2

The results of Study 1 indicate that the per capita GDP knowledge domain is suitable to investigate the processes underlying comparative judgments under uncertainty in general, and recognition-based inference in particular. Study 2 was designed to directly test the competing predictions of MaC and the RH that were laid out in the introduction.

Given that this is not the first study to measure RTs for comparative judgments involving RU pairs (e.g., Glöckner & Bröder, 2011; Hilbig & Pohl, 2009; Pachur & Hertwig, 2006; Volz et al., 2006), we also anticipated replicating a number of findings from previous studies. However, because most of the observed RT findings have been used as evidence *against* the RH, we focus, at this point, only on the one main RT effect that has been interpreted to be *consistent* with the RH. We postpone the review of other RT-related findings to the General Discussion, where we show how they relate to and are compatible with the MaC framework.

Pachur and Hertwig (2006; see also Volz et al., 2006) found that comparative judgments involving RU pairs are, on average, faster when the recognized pair member is chosen than when the unrecognized pair member is chosen. The authors attribute this finding to the "retrieval primacy" of recognition. That is, because recognition information is ready to use before any further (and potentially conflicting) knowledge is retrieved and entered into the inference process, responses that are solely based on recognition are, on average, faster than responses in which the retrieval of further knowledge triggers the suspension of the RH. However, there are at least two alternative explanations for this finding. One possibility is that the effect is an artifact due to subjective distance. Assuming that unknown pair members are inferred to have small criterion values and that choices are based on the comparison of subjective magnitudes, the subjective distance between pair members is, on average, most likely smaller in U-choice responses than in R-choice responses. According to this view, the effect should disappear once subjective distance is taken into account. A second possibility is that the effect simply reflects a response bias. Response biases are not uncommon in paired-comparison tasks (Banks, 1977; Vickers, 1979). Given that the frequency of R-choices is typically much higher than the frequency of U-choices in this task—sometimes reaching a ratio of 9:1 (Goldstein & Gigerenzer, 2002)—the threshold for R-choice responses might be set lower than the threshold for U-choice responses. This, in effect, would lead to faster choice times for the former compared with the latter. We explore these different accounts in the present study.

In order to test the competing predictions of MaC and the RH, it was necessary to obtain several secondary measures. These measures serve as important predictors in the analyses of choice times and choice patterns. In addition to the paired-comparison task, we included three further tasks in the study. (a) In a recognition task, participants were presented with individual country names and had to make timed recognition judgments. This task had two functions. First, the recognition judgments (in combination with the knowledge ratings, see below) served to classify country pairs in the comparison task into RU pairs and non-RU pairs. Second, the recognition judgment latencies were used as a control variable in the analysis of the paired-comparison data to isolate any potential effects related to recognition time. (b) The responses from a knowledge rating task, which was identical to Study 1, allowed us to remove any false alarms and misses from the responses in the recognition task. Therefore, the combination of recognition judgments and knowledge ratings enabled us to identify RU pairs whose members were consistently judged as recognized and unrecognized, respectively (Erdfelder et al., 2011; Pleskac, 2007). (c) Finally, in an estimation task, participants were asked to estimate the current per capita GDP of each country. The ranked per capita GDP estimates were used as a measure of subjective magnitude. The difference between two ranked estimates, in turn, served as a measure of the subjective difference between the per capita GDPs of two countries. For RU pairs, the subjective per capita GDP difference was computed as follows:

$$\text{Rank Difference}_{i,(r,u)} = \text{Ranked Estimate}_{i,r} - \text{Ranked Estimate}_{i,u}, \quad (1)$$

where i refers to the participant, r to the recognized country, and u to the unrecognized country.

The four tasks were administered in a fixed order: recognition task, paired-comparison task, knowledge rating task, and estimation task. In the paired-comparison task, we used a *repeated-set procedure* (Banks, 1977), analogous to Goldstein and Gigerenzer's (2002) original study on the RH. In a repeated-set procedure, a relatively small sample consisting of j items is drawn from a larger item population, and participants are presented with all $j(j-1)/2$ pairwise combinations. To maximize the representativeness of the sample of countries, we used four different lists of 16 countries in the paired-comparison task. Overall, these four lists included 32 different countries, which roughly corresponded to one fifth of the item population defined in Study 1.

Method

Participants. Two hundred ninety-six introductory psychology students took part in the study in exchange for partial course credit. Three participants were replaced by new ones because either extremely fast RTs or highly inconsistent responses across tasks implied that they were not properly attending to the task(s).

Materials. On the basis of the stimulus norms obtained in Study 1, 32 of the 153 countries were selected and grouped into four sets of eight countries (see Appendix A). Two sets (U1 and U2) consisted of countries that had relatively low recognition rates ($<52\%$), and two sets (R1 and R2) consisted of countries that had relatively high recognition rates ($>89\%$). The two sets of countries likely to be known were selected such that each included items with a wide range of subjective per capita GDPs, but actual values at or below the 67th percentile (Central Intelligence Agency, 2009). This, in effect, excluded any well-known prosperous countries from the stimulus materials. Given the above restrictions, each of the four sets included countries from as many different geographical regions as possible.

Each of the two sets of likely unknown countries was combined with each of the two sets of likely known countries, resulting in four 16-item lists (i.e., R1U1, R1U2, R2U1, R2U2). The four lists were rotated over participants, and the countries in an assigned list served as stimuli in the paired-comparison task. In the recognition judgment, knowledge rating, and estimation tasks, each participant was presented with all 32 countries, so that any countries that were not part of the paired-comparison task functioned as fillers.

Procedure. Each session consisted of four tasks—a recognition task, a paired-comparison task, a knowledge rating task, and an estimation task. In the recognition task, participants were instructed to make recognition judgments, as quickly and as accurately as possible, for 32 country names. Each trial was initiated by the participant with the spacebar. A fixation cross appeared in the center of the screen for 1,000 ms, which was subsequently overwritten by a country name. The country name remained on the screen until the participant pressed one of two response keys, *Z* or *M*. The assignment of response (*Recognize* vs. *Don't recognize*) to key (*Z* vs. *M*) was counterbalanced. The recognition task was preceded by a 20-trial practice block in which participants had to make recognition judgments for 20 actors and actresses (10 well-known and 10 little-known).

In the paired-comparison task, participants were presented with 120 country pairs (i.e., all pairwise combinations of the 16 countries in the respective list) and were instructed to choose, as quickly and accurately as possible, which of the two countries in each pair currently had the higher per capita GDP. Participants initiated each trial with the spacebar. A fixation cross appeared in the center of the screen for 1,000 ms followed by the presentation of two country names side by side. The response keys were *Z* (to select the country on the left) and *M* (to select the country on the right). The country names remained on the screen until a response was made. No feedback was provided about the accuracy of the response. The serial presentation order of country pairs and the order of countries within each pair (left vs. right) were randomized separately for each participant. The paired-comparison task was preceded by a 12-trial practice block in which the words *Richer* and *Poorer* were displayed side by side on the screen. Participants were instructed to press the response key (*Z* vs. *M*) that corresponded to the word *Richer* in each pair. If more than two errors were made during the practice trials, the entire practice block had to be repeated until this error criterion was met.

The knowledge rating task was identical to that of Study 1, with the exception that trials were self-paced rather than being initiated by the computer automatically. Participants provided knowledge ratings for all 32 countries.

In the final task, participants were asked to estimate the current per capita GDPs of the 32 countries to the nearest 100 Canadian dollars. The current value of the United States (C$60,900 at the time) was provided as a reference point in the instructions. Each trial was initiated with the spacebar, which triggered the presentation of a country name with an input field beneath it. Participants entered their response on the keyboard and confirmed it by pressing the *enter* key. The presentation order of countries in the recognition, knowledge rating, and estimation tasks was pseudorandomized separately for each participant and each task. This was done to ensure that likely known and likely unknown countries, as well as target and filler countries, were approximately equally distributed across the 32 trials.

Results and Discussion

Because several of the secondary measures serve as critical predictors in the analysis of the paired-comparison data, we briefly report the results of the recognition, knowledge rating, and estimation tasks first, before we continue with the main analyses of the choice and choice time data.

Secondary measures. Responses for countries that served as filler items in the recognition, knowledge rating, and estimation tasks were excluded from the analysis. Furthermore, we removed every item for which the recognition judgment and dichotomized knowledge rating were inconsistent (8.6% of responses). An inverse transformation (−1000/RT) was applied to the recognition latencies to reduce skewness. Visual inspection of the RT data also revealed a very small number of extreme outliers (0.2% of responses) that were excluded from further analyses. After data cleansing, we were left with complete secondary measures for a total of 4,322 trials.

Participants typically recognized about three fifths of the countries in the respective lists (*Mdn* = 62%, *IQR* = 20%). The by-item recognition rates had, as expected, a bimodal distribution. Sixteen countries had recognition rates above 90%, and the remaining ones had recognition rates below 59%. The median per capita GDP estimates for the recognized and unrecognized countries are shown in Figure 3. The graph reveals that the subjective per capita GDP values of unrecognized countries were very homogeneous and mostly fell into a relatively narrow interval between 10,000 and 15,000 Canadian dollars. This interval roughly corresponds to the low-to-medium portion of the distribution of estimated per capita GDP for the recognized countries. Thus, the present data, based on a sample of 32 countries, closely replicate the general pattern observed in Study 1.

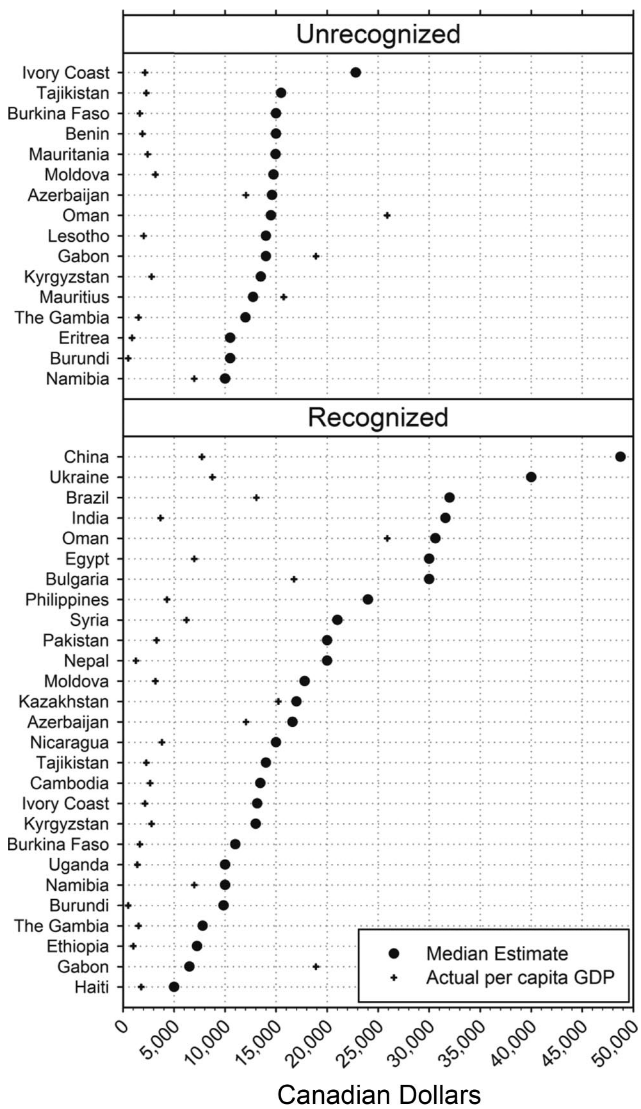

Figure 3. Actual and median estimated per capita gross domestic products (GDPs) for recognized and unrecognized countries (Study 2). Note that only data points with at least 15 observations are included in the graph.

Table 1 provides a summary of the interrelationship between all secondary measures—recognition times, knowledge ratings, and estimated and actual per capita GDPs. As in Study 1, the median by-participant rank-order correlation between estimated and actual per capita GDP was rather modest (*Mdn* r_S = .34, *IQR* r_S = .27), which again indicates that participants' specific knowledge of the criterion dimension was quite limited. Likewise, metric accuracy was poor (*Mdn OME* = 0.71, *IQR OME* = 0.34). It is also noteworthy that among all pairs of variables in Table 1, recognition times and knowledge ratings are correlated most highly (*Mdn* r_S = −.57, *IQR* r_S = .29), which should not be surprising as both have been used as measures of familiarity in the past (e.g., Brown & Siegler, 1992; Hertwig et al., 2008).

Choice times. All RU pairs were extracted from the paired-comparison data for which complete sets of valid secondary measures were available. Because seven participants recognized all countries presented in the paired-comparison task, we analyzed the data from 289 participants. An inverse transformation (−1000/RT) was used to reduce the skewness of the choice times. Furthermore, the inspection of by-subject quantile-quantile plots for the transformed choice times revealed a small number of extreme outliers that were removed (less than 0.3% of the data). This left us with responses for 408 unique RU pairs and 13,704 comparison trials overall. One of the critical predictors in the analysis of the paired-comparison data is the subjective distance in per capita GDP between the pair members. Therefore, we computed the rank difference for each RU pair, as defined by Equation 1, and took its absolute value. Because 16 countries were used in the paired-comparison task, this new variable, Absolute Rank Difference, had values ranging from 0 to 15. The greater the value of Absolute Rank Difference, the greater the subjective difference in per capita GDP between the two countries in a pair.[7] The item-based mean choice times are plotted as a function of mean absolute rank difference in Panel A of Figure 4. The graph reveals the characteristic RT function of the SDE, namely the inverse relationship between choice time and subjective distance on the criterion.

We analyzed the choice times using linear mixed-effects (LME) models, which incorporated subjects and items (RU pairs) as partially crossed random effects (Baayen, Davidson, & Bates, 2008; Bates, 2010; Pinheiro & Bates, 2000). This approach offered two advantages. (a) It enabled us to draw inferences about the effects of several covariates of interest, while simultaneously taking both subject and item variability into account. (b) Unlike many traditional approaches, LME modeling does not require the prior averaging over subjects or items, and thus allowed us to analyze trial-based responses, rather than subject or item means. All analyses were conducted in the R Environment (R Development Core

[7] Parallel analyses were conducted using the absolute difference of the ranked median estimates from the normative study (Study 1) as a distance measure. The obtained results are consistent with the ones reported here.

Table 1
Median By-Participant Spearman Rank-Order Correlations (and Interquartile Ranges) for Secondary Measures (Study 2)

Variable	1	2	3	4
1. Knowledge rating	—			
2. Recognition time	−.57 (.29)	—		
3. Estimated per capita GDP	.39 (.35)	−.27 (.37)	—	
4. Actual per capita GDP	.19 (.23)	−.10 (.35)	.34 (.27)	—

Note. GDP = gross domestic product.

Team, 2012) with the *lme4* (Bates, Maechler, & Bolker, 2011) and *languageR* (Baayen, 2011) packages. LME models were fitted by maximum likelihood, and p values and confidence intervals were estimated using Markov Chain Monte Carlo simulations.

We first fitted a preliminary LME model with transformed choice time as the response variable, Choice (R-choice vs. U-choice), standardized Recognition Time for the recognized (R) and unrecognized (U) country, and Trial as fixed effects, and Subject and RU Pair as random effects. This model served as the baseline model for subsequent model comparison. It incorporated predictors of choice time that, according to the RH, are either variables of interest or relevant control variables. The fixed-effects parameter estimates and fit statistics for this model are reported under Model 1 in Table 2.

Consistent with the results of previous studies (Pachur & Hertwig, 2006; Volz et al., 2006), Model 1 indicates that choices were made more quickly when the recognized country was chosen to have the higher per capita GDP than when the unrecognized country was chosen ($\hat{\beta} = 0.118$, $p < .001$). Furthermore, the three control variables in the model were also significant predictors of choice time. That is, choice time increased as the recognition time of the recognized country in a pair increased ($\hat{\beta} = 0.032$, $p < .001$). This result replicates the finding of a recent study by Marewski and Schooler (2011), who report the same relationship between the recognition time of the recognized pair member and choice time. Likewise, an increase in the recognition time of the unknown country in a pair was associated with an increase in choice time ($\hat{\beta} = 0.012$, $p < .001$). Finally, Model 1 indicates that participants became faster in responding to RU pairs as they progressed through the paired-comparison task ($\hat{\beta} = -0.003$, $p < .001$). We included the trial number as a covariate in order to isolate any practice effects that emerged in the course of the paired-comparison task. These types of effects are commonly observed in RT experiments (Baayen & Milin, 2010), and were expected in the present study because each country was presented across 15 different pairs.

To test the critical prediction of MaC, namely that choice time for R-choices decreases as a function of subjective distance, we fitted a second LME model to the data. In this model (Model 2 in Table 2), fixed-effects terms for Absolute Rank Difference and the interaction between Absolute Rank Difference and Choice were added. The interaction term was included to allow the distance effect to vary with choice. Furthermore, because the RH makes only predictions for R-choices, the competing predictions between MaC and the RH can only be tested for these responses. A likelihood ratio test indicated that the addition of the two terms significantly increased model fit, $\chi^2(2) = 113.22$, $p < .001$. Figure 5 shows partial-effects plots for the fixed-effects predictors in Model 2. The panels in the top row visualize the partial effects of Absolute Rank Difference, Choice, and their interaction: Choice times decreased as the value of Absolute Rank Difference increased ($\hat{\beta} = -0.010$, $p < .001$); however, the slope was somewhat shallower for U-choices than for R-choices ($\hat{\beta} = 0.005$, $p = .012$). This result indicates that, even after taking relevant control variables into account, comparative judgments involving RU pairs are characterized by the classic SDE, a finding consistent with the MaC account, but not the RH.

The somewhat shallower slope for U-choices is most likely related to the fact that there were relatively few RU pairs with large subjective distances where participants chose the unrecognized country as having the higher per capita GDP. To examine the interaction further, we analyzed the choice times for R-choices and U-choices separately. These LME analyses revealed that the effect of Absolute Rank Difference was significant in both R-choices ($\hat{\beta} = -0.009$, $p < .001$) and U-choices ($\hat{\beta} = -0.005$, $p = .002$),

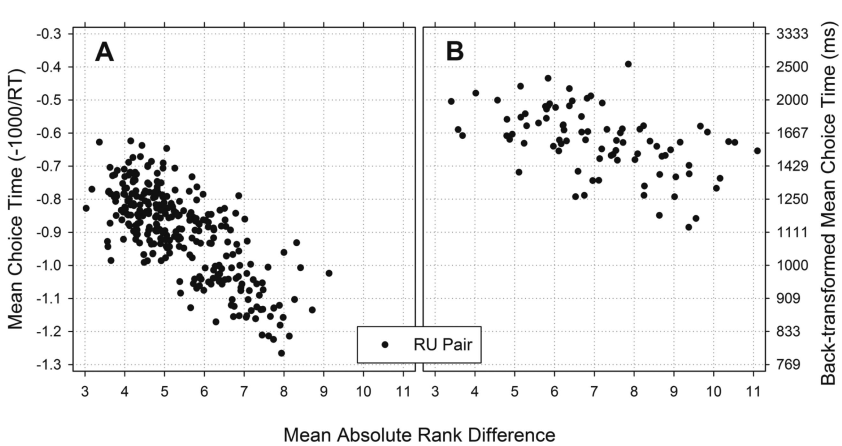

Figure 4. Mean choice time for RU pairs as a function of mean absolute rank difference (Panel A = Study 2, Panel B = Study 3). Note that the graph depicts only data points with at least 15 observations. RT = response time; RU pair = one recognized and one unrecognized item pair.

Table 2
Fit Statistics, Estimated Fixed-Effects Coefficients, and 95% Highest Posterior Density Intervals (HPDIs) From the LME Models Fitted to the RT Data (Study 2)

	Model 1		Model 2	
	Estimate	95% HPDI	Estimate	95% HPDI
Intercept	-0.782^{***}	[−0.807, −0.760]	-0.730^{***}	[−0.756, −0.705]
Trial	-0.003^{***}	[−0.003, −0.002]	-0.003^{***}	[−0.003, −0.002]
Choice (R-choice vs. U-choice)	0.118^{***}	[0.107, 0.134]	0.094^{***}	[0.072, 0.118]
Standardized Recognition Time (R)	0.032^{***}	[0.028, 0.043]	0.030^{***}	[0.026, 0.041]
Standardized Recognition Time (U)	0.012^{**}	[0.007, 0.023]	0.013^{**}	[0.008, 0.023]
Absolute Rank Difference			-0.010^{***}	[−0.012, −0.008]
Absolute Rank Difference × Choice			0.005^{*}	[0.001, 0.009]
Fit				
−2 log likelihood	7366		7253	
AIC	7382		7273	
BIC	7442		7348	

Note. Treatment coding was used for the predictor *Choice* (R-choice = 0, U-choice = 1). LME = linear mixed-effects; RT = response time; R = recognized; U = unrecognized; AIC = Akaike's information criterion; BIC = Bayesian information criterion.
$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

even when relevant control variables (i.e., Recognition time [R], Recognition time [U], Trial) were taken into account. This suggests that distance effects, the hallmark of the magnitude comparison process, emerged when participants chose both the recognized country and the unrecognized country.

Finally, the effect plot of the interaction between Absolute Rank Difference and Choice in Figure 5 suggests that R-choices were made more quickly than U-choices across the entire range of rank difference values. Our interpretation of this finding is that it most likely reflects a response bias. As described in more detail below, the ratio of R-choices to U-choices in this study was, on average, about 3:1. Given the unequal response probabilities for R-choices and U-choices in the task, a bias against the U-choice response might be set up, such that the threshold for R-choice responses is generally set somewhat lower than the threshold for U-choice responses. This, in turn, leads to faster RTs for R-choices because fewer iterations are required to reach the response threshold.

Choice. The median by-participant proportion of R-choices was .78 (*IQR* = .22). (The proportions of R-choice responses for each participant are shown in Figure S1 in the Supplemental Materials.) On average, about three fifths of participants' judgments were correct (*Mdn* = .60, *IQR* = .13). For descriptive purposes, the by-item proportions of R-choices are plotted as a function of mean rank difference in Panel A of Figure 6. We used the rank difference as a measure of subjective distance (rather than the absolute rank difference) here because the direction of the

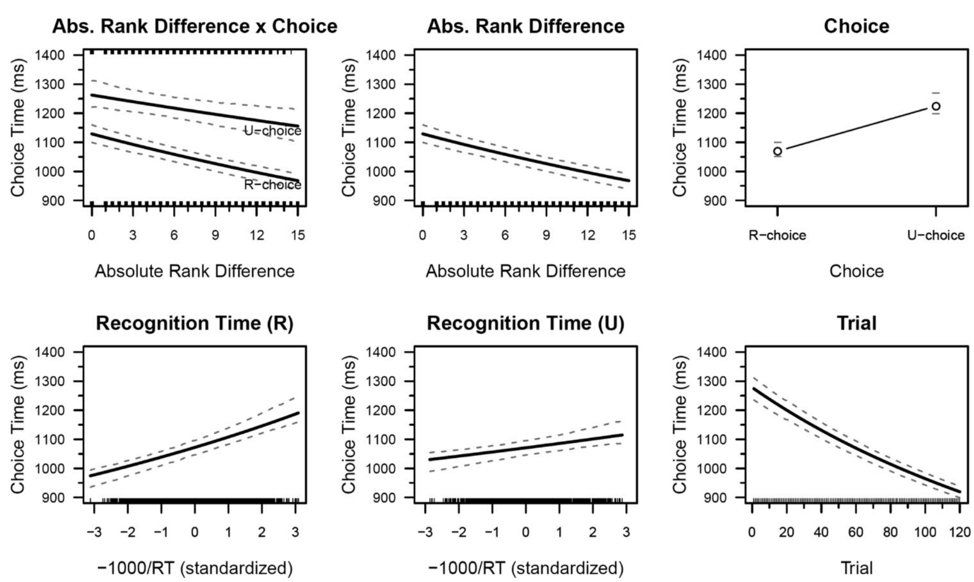

Figure 5. Partial-effects plots for the fixed-effects predictors in the linear mixed-effects model (Model 2) fitted to the choice times for one recognized and one unrecognized item pairs (Study 2). Dashed lines represent 95% highest posterior density intervals based on 10,000 samples from the posterior distribution of the parameters. Abs. = Absolute; R = recognized; U = unrecognized; RT = response time.

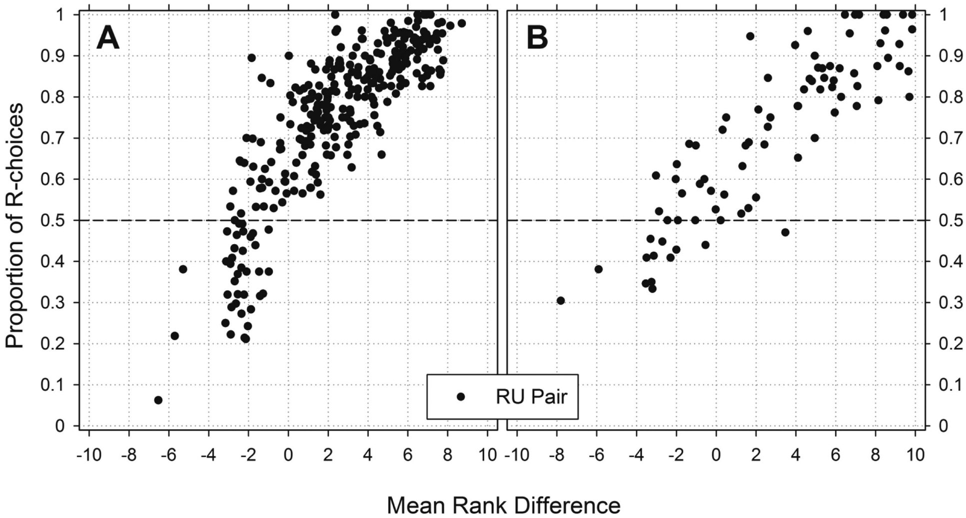

Figure 6. Proportion of R-choices as a function of mean rank difference (Panel A = Study 2, Panel B = Study 3). Only data points with at least 15 observations are shown. RU pair = one recognized and one unrecognized item pair.

difference is crucial for predicting choice behavior. The graph reveals a systematic relationship between the two variables. For positive values of rank difference (i.e., the recognized country is estimated to have the higher per capita GDP), R-choices are predominant, and their frequency increases as the value of rank difference becomes more positive. Conversely, for negative values of rank difference (i.e., the unrecognized country is estimated to have the higher per capita GDP), the opposite pattern is seen. The choice data were analyzed using logistic mixed-effects models[8] (Agresti, 2002; Jaeger, 2008). Parallel to the choice time analysis, all models incorporated Subject and RU Pair as random effects. The fit statistics and fixed-effects coefficients of the models are reported in Table S2 in the Supplemental Materials.

An initial model that included the control variables Recognition Time (R), Recognition Time (U), and Trial as fixed effects did not fit better than a model without them, $\chi^2(3) = 3.75$, $p = .290$, and the three control variables were therefore removed from subsequent models. The nonsignificant coefficient for Recognition Time (R) ($\hat{\beta} = -0.050$, $z = -1.46$, $p = .144$) is worth noting because it suggests that one of the proposed suspension criteria of the RH, the low retrieval fluency of the recognized pair member (Marewski, Gaissmaier, et al., 2010), does not account for the choice patterns observed in the present study. In contrast, a model that incorporated Rank Difference as a fixed effect (Model 2 in Table S2) improved model fit relative to the intercept-only model (Model 1 in Table S2), $\chi^2(1) = 963.49$, $p < .001$. For every one-unit increase in Rank Difference, the log odds of R-choice increased by 0.171. This finding is consistent with the idea that people relied on a magnitude comparison process, but is once again at odds with the predictions of the RH.

Taken together, the RT and choice analyses provide converging evidence for the MaC account as they demonstrate that comparative judgments involving RU pairs are, like many other types of comparative magnitude judgments, characterized by distance effects. In contrast, it is problematic to find a parsimonious account of this effect if one assumes that participants relied mostly on the cue-based RH, considering in particular that we controlled for several potential confounding variables. In Study 3, we examine whether and to what extent the observed distance effect is related to the repeated-set procedure of the paired-comparison task.

Study 3

In Study 2, we used a repeated-set procedure, as has been the case in most prior research on the RH (but see Hertwig et al., 2008; Marewski & Schooler, 2011; Volz et al., 2006). That is, participants were presented with all possible pairs created from a relatively small set of items drawn from a larger item population. However, this procedure does have an obvious drawback (Banks, 1977; Bower, 1971). Because the same items are repeatedly presented across different pairs, it is possible that, during the course of the experiment, people create ad hoc cognitive structures that represent the linear ordering of the items used in the paired-comparison task. As a result, participants might rely predominantly on these "temporary data sets" in their comparative judgments, instead of retrieving information from semantic memory anew on each trial. To remedy this problem, researchers in the 1970s sometimes used an *infinite-set procedure* (e.g., Banks & Flora, 1977; Pavio, 1975). That is, they used a relatively large set of items and created unique binary combinations such that each item was presented in only one pair during the comparison task. We use the same approach in the present study in order to exclude the possibility that the observed distance effect in Study 2 is an artifact of the repeated-set procedure.

A second advantage of the infinite-set procedure is that it provides a more "naturalistic" context for recognition-based inferences. This is because the procedure ensures that, at the point of

[8] Logistic mixed-effects models were fitted by Laplace approximation. *Z* tests were used to test the statistical significance of individual fixed-effects coefficients. The 95% confidence intervals for these coefficients are also *z*-based.

the comparative judgment, a truly unknown pair member is encountered for the first time. In other words, the infinite-set procedure eliminates the artifact of the repeated-set procedure in which, due to multiple presentations, unrecognized items can become familiar objects during the course of the experiment. This, in effect, can blur the line between a genuine RU pair and an RR pair.

In order to address the issues of linear orderings and item repetition, we made two major changes to the design of the previous study. First, the sequence of the tasks was modified so that the paired-comparison task preceded both the knowledge rating and estimation tasks. This ensured that all country names were presented to participants for the first time in the study when the comparative judgment had to be made. However, because this modification made it impossible to measure "clean" recognition times, we did not include a recognition task. Instead, we used the mean recognition times from the previous study as covariates in the statistical analysis. Second, we used an infinite-set procedure in the paired-comparison task.

Method

Participants. Three hundred three undergraduate psychology students participated in the study in exchange for partial course credit. None of them had taken part in the previous studies.

Materials. Each participant was presented with a total of 80 different countries. Twenty of the 80 countries were the same for all participants (fixed set) and had been used in the previous study. They consisted of 10 likely known countries and 10 likely unknown countries (see Appendix B). The other 60 countries (random set) were randomly sampled for each participant from the remaining countries of the item population defined in Study 1. The 10 likely known and 10 likely unknown countries from the fixed set were randomly paired for each participant, resulting in 10 likely RU pairs. Therefore, across participants, we expected to obtain responses to 100 unique RU pairs. The 60 countries from the random set served as filler items in all three tasks. Pairs of filler countries were created randomly.

Procedure. The experimental session consisted of a paired-comparison task, a knowledge rating task, and an estimation task. The procedures within each task were the same as in Study 2, with a few minor exceptions. (a) The 40 trials of the paired-comparison task were pseudorandomized separately for each participant. During the first 10 trials, only filler pairs were presented. Then one expected RU pair (from the fixed set) was presented at a random position in every successive trial triplet. (b) The practice phase that preceded the paired-comparison task was extended to 20 trials. (c) The trial order in the knowledge rating and estimation tasks was randomized (rather than pseudorandomized) separately for each participant and task.

Results and Discussion

We used the same data-analytic strategy as in Study 2. All reported analyses are restricted to the 20 countries from the fixed set.

Secondary measures. We computed recognition rates, based on the dichotomized knowledge ratings, for both subjects and items. On average, participants recognized about three fifths of the countries from the fixed set (Mdn = 60%, IQR = 18%). The recognition rates for the 20 countries were, as before, bimodal. Ten countries had recognition rates of less than 51%, and the other 10 had recognition rates greater than 95%. The median per capita GDP estimates for the recognized and unrecognized countries are shown in Figure 7. As in Study 2, the median estimates for the unrecognized countries were quite uniform, and all fell into the low-to-medium range of subjective per capita GDP. The accuracy of the estimates was again relatively low, both on a metric level ($Mdn\ OME$ = 0.71, $IQR\ OME$ = 0.37) and on a mapping level ($Mdn\ r_S$ = .34, $IQR\ r_S$ = .29). Once more, this suggests that participants' specific knowledge of the per capita GDPs of these countries was rather limited.

Choice times. From all comparison trials involving the 100 expected RU pairs, we selected the ones for which participants' knowledge ratings indicated that they were actual RU pairs. An inverse transformation (−1000/RT) was applied to the choice times to reduce skewness and the influence of outliers. Seven participants recognized all of the 20 countries in the fixed set, which left us with data from 296 participants and a total of 2,109 comparison trials. Furthermore, we computed several predictor variables. (a) As in Study 2, we used the absolute value of the rank difference defined by Equation 1 as an index of the subjective per capita GDP difference between two countries. Because there were 20 countries in the fixed set, the values of the predictor variable Absolute Rank Difference in this study ranged from 0 to 19. (b) To control for the effects of recognition time, we computed two variables: Mean Recognition Time (R) and Mean Recognition Time (U). The values of these variables consisted of the mean-transformed RTs for the recognized and unrecognized countries, respectively, that were obtained in the recognition judgment task of Study 2 (mean RTs were conditioned on the recognition judgment). Both variables were standardized.

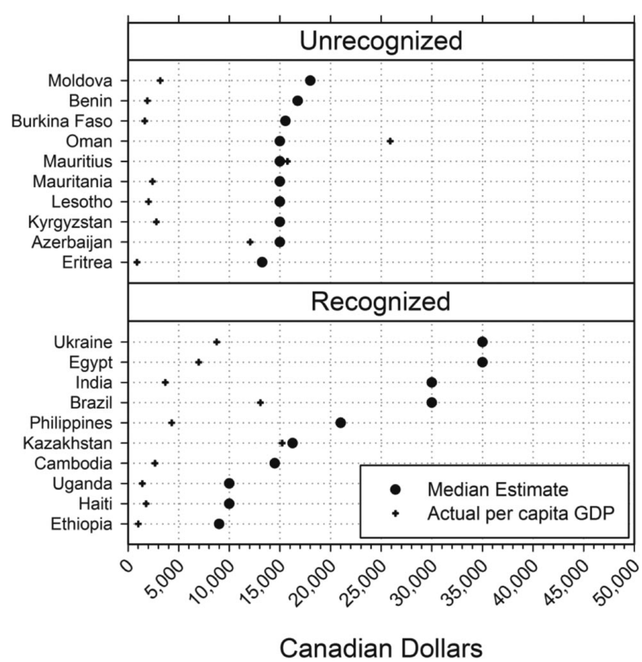

Figure 7. Actual and median estimated per capita gross domestic products (GDPs) for the recognized and unrecognized countries from the fixed set (Study 3).

The item-based mean choice times are plotted as a function of mean absolute rank difference in Panel B of Figure 4. The graph reveals that the average choice times were, in general, slower than in Study 2. Yet, the inverse relationship between choice time and subjective distance emerges again.

Following the same logic as in the previous study, we began the analysis by fitting an initial LME model to the RT data (see Model 1 in Table 3), which included predictors that, according to the RH, should be relevant either as variables of interest or as control variables. Model 1 incorporated Choice (R-choice vs. U-choice), Mean Recognition Time (R), Mean Recognition Time (U), and Trial as fixed effects, and Subject and RU Pair as random effects. The fixed-effects coefficients in Table 3 indicate that choices were made more quickly when the recognized country was chosen to have the higher per capita GDP than when the unrecognized country was chosen ($\hat{\beta} = 0.059$, $p < .001$). Furthermore, the speed of the choices was related to the mean recognition times of the individual pair members. That is, choice time increased as both the mean recognition time for the recognized pair member ($\hat{\beta} = 0.059$, $p < .001$) and the unrecognized pair member ($\hat{\beta} = 0.015$, $p = .009$) increased. Both of these results replicate the findings from Study 2. Finally, Model 1 suggests that the decrease in choice time during the course of the comparison task was much smaller in the present study ($\hat{\beta} = -0.001$, $p = .069$). This result, as well as the longer choice times in general, is most likely a consequence of the infinite-set procedure in which every pair consists of two previously unpresented items.

In a second LME model, we also included Absolute Rank Difference and the interaction between Absolute Rank Difference and Choice as fixed effects (see Model 2 in Table 3). The addition of these two terms improved model fit significantly, $\chi^2(2) = 16.94$, $p < .001$. Figure 8 depicts the partial effects of the fixed-effects predictors included in Model 2. The plots show that choice time decreased as the value of Absolute Rank Difference increased ($\hat{\beta} = -0.005$, $p < .001$) but that the slope for U-choices was shallower than the slope for R-choices ($\hat{\beta} = 0.004$, $p = .042$). Separate LME analyses for R-choices and U-choices revealed that Absolute Rank Difference was a significant predictor of transformed choice time when the recognized country was chosen ($\hat{\beta} = -0.004$, $p < .001$), but not when the unrecognized country was chosen ($\hat{\beta} = -0.001$, $p = .943$). (These models also included Mean Recognition Time [R], Mean Recognition Time [U], and Trial as control variables.) Because the distance effect for R-choices constitutes the critical differential prediction between the MaC and RH accounts, its replication in the present study provides converging evidence for MaC. Furthermore, it suggests that the distance effect observed in Study 2 is not simply an artifact of the repeated-set procedure, because the effect also emerges when the possibility for creating linear orderings is eliminated. In contrast to Study 2, we did not find a distance effect for U-choice responses. However, it is worth noting that, due to the infinite-set procedure, the number of data points for U-choices was much smaller in this study ($n = 610$) than in the previous one ($n = 3{,}364$). Therefore, it is difficult to draw conclusions from the obtained null result for U-choices, and we leave it to future studies to systematically explore the factors related to the emergence of distance effects in U-choices.

Choice. On average, participants chose the recognized country in the critical RU pairs three quarters of the time (*Mdn* = .75, *IQR* = .31; see also Figure S2 in the Supplemental Materials). The median by-participant proportion of correct judgments was .60 (*IQR* = .25). To visualize the distance effect, the by-item proportions of R-choices are plotted as a function of mean rank difference in Panel B of Figure 6.

In an initial logistic mixed-effects model, we included the control variables Mean Recognition Time (R), Mean Recognition Time (U), and Trial as fixed effects, and Subject and RU Pair as random effects. Because the coefficient for Mean Recognition Time (U) was nonsignificant, we removed this predictor and refitted the model. The reduced model (see Model 1 in Table S3) indicates that participants became less likely to choose the recognized country in an RU pair as the value of Mean Recognition Time (R) increased ($\hat{\beta} = -0.680$, $z = -6.42$, $p < .001$). This result, which differs from that of Study 2, is most likely related to the fact that we used averaged recognition times as a covariate in the analysis rather than subject-specific ones. We return to the

Table 3
Fit Statistics, Estimated Fixed-Effects Coefficients, and 95% Highest Posterior Density Intervals (HPDIs) From the LME Models Fitted to the RT Data (Study 3)

	Model 1		Model 2	
	Estimate	95% HPDI	Estimate	95% HPDI
Intercept	-0.622^{***}	[−0.661, −0.594]	-0.585^{***}	[−0.626, −0.547]
Trial	$-0.001^{\dagger}$	[−0.002, 0.000]	$-0.001^{\dagger}$	[−0.002, 0.000]
Choice (R-choice vs. U-choice)	0.059^{***}	[0.052, 0.097]	0.033^{*}	[0.004, 0.080]
Standardized Mean Recognition Time (R)	0.059^{***}	[0.046, 0.069]	0.055^{***}	[0.042, 0.065]
Standardized Mean Recognition Time (U)	0.015^{**}	[0.004, 0.026]	0.015^{**}	[0.004, 0.025]
Absolute Rank Difference			-0.005^{***}	[−0.008, −0.003]
Absolute Rank Difference × Choice			0.004^{*}	[0.000, 0.009]
Fit				
−2 log likelihood	−329.5		−346.4	
AIC	−313.5		−326.4	
BIC	−268.2		−269.9	

Note. Treatment coding was used for the predictor *Choice* (R-choice = 0, U-choice = 1). LME = linear mixed-effects; RT = response time; R = recognized; U = unrecognized; AIC = Akaike's information criterion; BIC = Bayesian information criterion.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

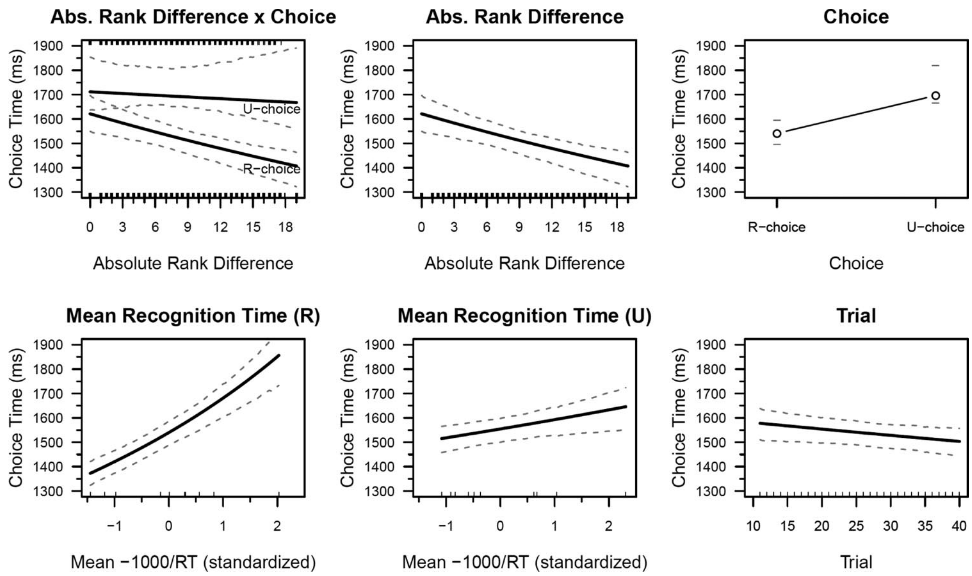

Figure 8. Partial-effects plots for the fixed-effects predictors in the linear mixed-effects model (Model 2) fitted to the choice times for one recognized and one unrecognized item pairs (Study 3). Dashed lines represent 95% highest posterior density intervals. Abs. = Absolute; R = recognized; U = unrecognized; RT = response time.

relationship between recognition time and the proportion of R-choices in the General Discussion. Model 1 also indicates that participants in Study 3 became somewhat more likely to choose the recognized country in a pair as they progressed through the paired-comparison task ($\hat{\beta} = 0.022$, $z = 3.35$, $p < .001$). Critically, a model (see Model 2 in Table S3) that included Rank Difference as a predictor improved model fit relative to Model 1, $\chi^2(1) = 175.47$, $p < .001$. A one-unit increase in Rank Difference increased the log odds of R-choice by 0.117. Again, this result is consistent with the MaC account, but not the RH.

In sum, the results of Study 3 provide converging evidence for the MaC framework. Critically, we replicated the SDE in R-choices when we changed the procedure from a repeated-set to an infinite-set one. This suggests that the distance effect observed in Study 2 is not simply due to participants setting up and relying on linear orderings. Instead, the fact that the distance effect also emerged, at least in R-choices, when participants always had to compare two previously unpresented countries, indicates that it is a robust phenomenon.

General Discussion

Over the last several decades, an enormous amount of effort has been devoted to the study of comparative magnitude judgments. Yet, this research has led to two independent and fundamentally different theoretical accounts of the nature of the underlying cognitive processes. The studies reported in this article were designed to directly test competing predictions derived from these two theoretical perspectives—the FFH framework, on the one hand, and the MaC framework, on the other. In particular, we focused on comparative judgments involving one recognized and one unrecognized object in a knowledge domain, national per capita GDP, in which recognition can serve as a probabilistic cue. In two paired-comparison studies, we replicated the common finding that people frequently infer that the recognized object in an RU pair has the higher criterion value (Goldstein & Gigerenzer, 2002); however, we also found that both choices and choice times are systematically related to the subjective magnitude difference between the compared objects. This distance effect served as the critical differential prediction, as it is the classic hallmark of the magnitude comparison process, but does not directly follow from a cue-based strategy such as the RH.

Below, we discuss the implications of these findings in three separate sections. First, we propose that MaC can serve as a general framework for understanding comparative magnitude judgments under (un)certainty. Second, we illustrate how central findings reported in the RH literature relate to and can be accounted for by MaC. Third, we discuss the relationship between the FFH and MaC frameworks, as well as the implications for future research.

MaC as a General Framework for Understanding Comparative Magnitude Judgments Under (Un)certainty

The core assumption of the MaC framework is that an iterative two-stage magnitude comparison process plays a central role in comparative magnitude judgments under certainty *and* under uncertainty. That is, magnitude comparisons that are structurally the same but vary along the certainty–uncertainty continuum differ in terms of the processes involved in the initial magnitude-generation stage, but not in terms of the comparison process itself. According to this view, the same general processing architecture is commonly used across different choice situations and environments.

This conceptualization differs from the FFH framework in two important ways. First, the FFH framework, more specifically, the theory of probabilistic mental models (Gigerenzer et al., 1991), holds that comparative magnitude judgments under certainty and under uncertainty involve separate sets of strategies that require fundamentally different cognitive processing architectures (i.e., LMMs vs. PMMs). In fact, as the number of simple heuristics has grown over the years, the "toolbox" is now assumed to include strategies that are applicable in the same choice situations and environments, yet involve fundamentally different underlying cognitive mechanisms. For example, take-the-best (Gigerenzer & Goldstein, 1996) and the fluency heuristic (Schooler & Hertwig, 2005) are both applicable in comparisons of two recognized objects. Yet, whereas the former is based on a serial cue-matching process, the latter involves the comparison of relative retrieval time.

Second, an integral part of the FFH framework is the idea that people choose among different strategies in comparative judgments under uncertainty as they are confronted with different choice situations and environments (Marewski & Schooler, 2011; Rieskamp & Otto, 2006). That is, the RH might be used in situations with one recognized and one unrecognized object, but the fluency heuristic, take-the-best, and other heuristics are applied when two recognized objects are compared. Although we do not reject the idea of multiple strategies playing a role in this inference task, we question why people should switch, in RH-relevant contexts even by default, to a fundamentally different set of strategies in judgments under uncertainty, when they already have a mechanism in place that is widely applied in judgments under certainty. It should be evident that if people were to also rely on the MaC process in comparative judgments under uncertainty, then this would simplify or even eliminate the strategy selection problem, a problem that is inherent in the FFH framework (Newell, 2005). This, and the fact that MaC incorporates the same magnitude-generation processes used in other magnitude judgment tasks (e.g., estimation), makes the account particularly parsimonious.

The empirical evidence for the relevance of the MaC process in comparative magnitude judgments under uncertainty extends beyond the present findings. The results of a recent study by Brown and Tan (2011) suggest that the MaC process also plays an important role in the "standard" choice situation involving two recognized objects. In their study, participants were asked to decide which of two automobiles was more expensive, and the obtained RT and choice measures were used to test competing predictions between MaC and the take-the-best heuristic. Whereas take-the-best predicted that choice times would increase with the number of necessary cue evaluations (Bröder & Gaissmaier, 2007), MaC predicted that choice times would primarily vary as a function of subjective distance. Even though predictions for many pair types overlapped, it was possible to formulate competing predictions for a subset of them. The obtained RT (and choice) patterns for these critical pair types were largely consistent with the MaC predictions. This result, in combination with the present findings, thus supports the view that the same type of process, namely MaC, is used in the present inference task across different choice situations and environments.

Furthermore, similar RT and choice functions as the ones predicted by MaC have been reported in studies on the fluency heuristic (Hertwig et al., 2008; Marewski & Schooler, 2011). This research shows that choices and choice times in comparisons of two recognized objects tend to vary as a function of the difference in retrieval time between the compared objects. That is, as the difference in retrieval time increases, choice time decreases, and the object with the faster retrieval time is chosen more frequently. This pattern is typically interpreted as evidence for the fluency heuristic, which states that one should infer that the more fluently retrieved pair member has the higher criterion value (Schooler & Hertwig, 2005). The aforementioned studies did not include measures of subjective magnitude, and therefore it is difficult to evaluate the role of the MaC process. However, the direct relationship between MaC and the fluency heuristic should be evident. Unlike many knowledge-based heuristics (e.g., take-the-best), the fluency heuristic is actually based on a comparison rather than a cue evaluation process. And, in fact, the heuristic can be considered a special case of MaC, namely, one where magnitude generation is entirely based on item familiarity.

In an attempt to integrate the metrics-and-mappings framework for understanding real-world quantitative estimation (Brown & Siegler, 1993) with the magnitude comparison literature, we specified a number of candidate processes that likely play a role in the initial stage of the MaC process. In general, if magnitude values cannot be directly retrieved from memory, magnitudes are generated on the basis of two sources of information: familiarity-based intuitions and task-relevant domain-specific knowledge. The relative weighting of these two sources of information depends on their ecological validities. Research on real-world estimation suggests that people primarily rely on familiarity-based intuitions when domain-specific knowledge is sparse and that they weight knowledge more heavily when it is available and task-relevant (Brown, 2002; Brown & Siegler, 1993). A similar pattern seems to emerge in paired-comparison tasks. In a recent study, Marewski and Schooler (2011) concluded that the fluency heuristic is most applicable and frequently used in comparative judgments when item knowledge is limited or absent but that the heuristic is outperformed by knowledge-based strategies when knowledge is available. The authors account for this pattern by assuming that people switch between familiarity-based and knowledge-based heuristics, given the choice situation and environment. However, according to MaC, this pattern emerges because familiarity and knowledge are differentially weighted in the magnitude-generation process. Thus, in cases in which magnitude generation is entirely based on item familiarity, MaC effectively reduces to the fluency heuristic.

A final issue concerns the representation and use of real-world knowledge. According to the most prominent knowledge-based heuristic, take-the-best (Gigerenzer & Goldstein, 1996), information retrieval is assumed to consist of a validity-guided serial search through a database of probabilistic cues. Although several theoretical problems have been identified with this assumption (Dougherty et al., 2008), the general view is still widely held both by proponents and critics of simple heuristics. However, besides its theoretical problems, this cue-centric view is also detached from and largely inconsistent with a body of research on plausible reasoning strategies (e.g., Brown, 2002; Collins & Michalski, 1989; Collins et al., 1975; A. Friedman & Brown, 2000a; W. J. Friedman, 1993; Graesser & Franklin, 1990; Reder, 1987). This line of research suggests that people generally take a much more opportunistic approach when answering questions about real-world domains. In broad terms, this approach holds that information search starts with the key elements of the question and sets off a retrieval-inference cycle. The retrieval of a given fact depends on how strongly it is associated with those key elements. The most accessible information is retrieved first. If one or multiple facts can be retrieved, particular inference patterns are triggered. The specific types of inferences that are drawn during this process are qualified by the nature of the retrieved information. Collins and Michalski (1989) have demonstrated that it is possible to develop a formal representation of many plausible inference patterns used in answering everyday questions, and this approach could provide a realistic and implementable alternative to the PMM architecture.

MaC Account of Other Findings in the RH Literature

In this section, we illustrate, on a more specific level, how the MaC framework can account for several pertinent findings reported in the RH literature. As most previous investigations of this choice context have not included measures of subjective magnitude (but see Hilbig et al., 2009; Pachur & Hertwig, 2006), we cannot evaluate MaC directly. However, by addressing these additional findings, we hope to illustrate the explanatory power of MaC and, at the same time, lay out a set of predictions that might serve as the basis for future research.

First, in attempts to falsify the hypothesis of noncompensatory use of the recognition cue, several researchers have tried to demonstrate that different types of RU pairs elicit different RT and choice patterns. One such difference concerns RU pairs in which the recognized pair member is merely recognized versus those in which further knowledge is available (e.g., Hilbig & Pohl, 2009; Pohl, 2006). In general, it has been found that choices are made more quickly and R-choices are more common for the latter than the former type of RU pair. According to MaC, this result is related to the fact that the average subjective distance between pair members in "mere recognition" RU pairs is likely smaller than the average subjective distance in "further knowledge" RU pairs. Consequently, average choice times should be faster and average rates of R-choice higher for the latter compared with the former type of RU pair.

Second, another common finding in the RH literature is that the proportion of R-choices is inversely related to the recognition time of the recognized pair member (e.g., Marewski & Schooler, 2011; Newell & Fernandez, 2006). We replicated this result in Study 3 when we used the mean recognition times for the recognized countries from Study 2 as a covariate in the analysis of choice behavior. We did this in order to rule out the alternative explanation that the observed relationship between the proportion of R-choices and subjective distance was related to the suspension of RH in RU pairs in which the recognized country had a relatively slow retrieval time (Marewski, Gaissmaier, et al., 2010).

From the MaC point of view, the systematic relationship between retrieval time and the proportion of R-choices observed in previous studies could be due to (a) people primarily relying on familiarity in the magnitude-generation process and/or (b) the fact that, for a given item, the degree of familiarity and the amount of task-relevant knowledge are generally correlated. As described above, the generation of magnitudes is based on a weighted blend of familiarity-based intuitions and domain-specific knowledge. Therefore, in domains in which people primarily rely on familiarity to generate magnitudes, familiar items (which tend to have faster recognition times) will be inferred to have relatively large magnitudes and are therefore more likely to be chosen as having the higher criterion value. Furthermore, in studies in which the effects of familiarity and knowledge were not disentangled, the reported pattern could have also been due to the correlation between familiarity and knowledge. That is, people are more likely to have task-relevant knowledge for familiar items, and these items, on average, tend to have higher criterion values.

A third set of findings pertains to differences in the average proportion of R-choices in comparison tasks involving samples of items drawn from different knowledge domains or different reference classes (Gigerenzer & Goldstein, 2011; Pohl, 2006). According to the RH, these differences mainly emerge because people use the value of the recognition validity, which has to be estimated on the basis of the specific reference class, as a criterion to evaluate whether or not to apply the RH. Thus, the higher the recognition validity, the higher the average rate of R-choices should be. In a recent meta-analysis, Gigerenzer and Goldstein (2011) reported a Pearson correlation of .57 between the recognition validities and mean proportions of R-choices in a set of 43 studies on the RH and concluded that "people tend to rely on the recognition heuristic consistently when the validity for them is high, but when the validity decreases, people increasingly suspend the default and follow some other strategy" (p. 105).

In contrast, the MaC view holds that differences in the average rate of R-choices are related to the relative overlap of the subjective magnitude distributions for the recognized and unrecognized items. For example, in the present study, we found that participants typically inferred that unknown countries have low-to-medium per capita GDPs, and the average proportion of R-choices was generally about 75%. However, on a different criterion dimension, the relative overlap of the subjective magnitude distributions for recognized and unrecognized countries might be smaller, which, given a representative sample, should lead to higher average proportions of R-choices in the comparison task. In fact, we have conducted a parallel set of studies in the knowledge domain of country population (Schweickart, Brown, & Lee, 2009). Here, people commonly infer that unknown countries have very small populations, and we found that in comparison tasks, the average proportion of R-choices is generally higher than the one observed in the per capita GDP domain. Furthermore, Pohl's (2006) finding that the average proportion of R-choices drops to chance levels in domains in which recognition is unrelated to the criterion can be accounted for in a similar way. In these cases, the subjective magnitudes for the unrecognized items might either be randomly distributed across the target dimension or be consistently located at the center of the distribution for the recognized items, thus leading to an average proportion of R-choices of about 50%.

In sum, the MaC framework provides a parsimonious account of major findings reported in the RH literature, without having to resort to restrictive usage conditions or an elaborate catalogue of suspension criteria.

Conclusions and Implications for Future Research

In the late 1970s, leading researchers on the topic of memory-based comparative judgments concluded that "the generality of the distance effect suggests that it may represent one of those critical phenomena that reflect basic psychological principles" (Potts et al., 1978, p. 243). More than 30 years later, we provide evidence consistent with this notion by showing that distance effects are also a characteristic of the special case of comparative magnitude judgments under uncertainty that involve one recognized and one unrecognized object. Thus, our finding fits seamlessly into more than a century of research documenting distance effects in a wide variety of comparison tasks.

The research presented in this article indicates that the MaC process is commonly used in the task at hand. Nonetheless, it is still possible that people, under certain conditions, might use simple heuristics to make comparative judgments. Of course, the question of when, how, and why simple heuristics are adopted is an empirical one. However, it seems clear that research directed toward understanding this issue must take the MaC process into account and acknowledge its potential relevance. More generally, we suspect that the two theoretical frameworks, MaC and FFH, can complement one another and that a full description of the cognitive mechanisms underlying comparative magnitude judgments under uncertainty will be one that integrates both approaches.

References

Agresti, A. (2002). *Categorical data analysis*. Hoboken, NJ: Wiley. doi:10.1002/0471249688

Baayen, R. H. (2011). languageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics" (Version 1.4) [R package]. Retrieved from http://cran.r-project.org/web/packages/languageR/index.html

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research, 3,* 12–28.

Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11, pp. 101–159). New York, NY: Academic Press. doi:10.1016/S0079-7421(08)60476-4

Banks, W. P., & Flora, J. (1977). Semantic and perceptual processes in symbolic comparisons. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 278–290. doi:10.1037/0096-1523.3.2.278

Banks, W. P., Fujii, M., & Kayra-Stuart, F. (1976). Semantic congruity effects in comparative judgments of magnitudes and digits. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 435–447. doi:10.1037/0096-1523.2.3.435

Bates, D. M. (2010). lme4: Mixed-effects modeling with R [Prepublication version]. New York, NY: Springer. Retrieved from http://lme4.r-forge.r-project.org/book/

Bates, D. M., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes (Version 0.999999–0) [R package]. Retrieved from http://lme4.r-forge.r-project.org/

Beach, L. R., & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy of Management Review, 3,* 439–449. doi:10.2307/257535

Birnbaum, M. H., & Jou, J. (1990). A theory of comparative response times and "difference" judgments. *Cognitive Psychology, 22,* 184–210. doi:10.1016/0010-0285(90)90015-V

Bower, G. H. (1971). Adaptation-level coding of stimuli and serial position effects. In M. H. Appley (Ed.), *Adaptation-level theory* (pp. 175–201). New York, NY: Academic Press.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science, 282,* 746–749. doi:10.1126/science.282.5389.746

Bröder, A., & Eichler, A. (2006). The use of recognition information and additional cues in inferences from memory. *Acta Psychologica, 121,* 275–284. doi:10.1016/j.actpsy.2005.07.001

Bröder, A., & Gaissmaier, W. (2007). Sequential processing of cues in memory-based multiattribute decisions. *Psychonomic Bulletin & Review, 14,* 895–900. doi:10.3758/BF03194118

Brown, N. R. (2002). Real-world estimation: Estimation modes and seeding effects. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 41, pp.321–359). New York, NY: Academic Press. doi:10.1016/S0079-7421(02)80011-1

Brown, N. R., Cui, X., & Gordon, R. D. (2002). Estimating national populations: Cross-cultural differences and availability effects. *Applied Cognitive Psychology, 16,* 811–827. doi:10.1002/acp.830

Brown, N. R., Rips, L. J., & Shevell, S. K. (1985). The subjective dates of natural events in very-long-term memory. *Cognitive Psychology, 17,* 139–177. doi:10.1016/0010-0285(85)90006-4

Brown, N. R., & Siegler, R. S. (1991). Subjective organization of U.S. presidents. *The American Journal of Psychology, 104,* 1–33. doi:10.2307/1422849

Brown, N. R., & Siegler, R. S. (1992). The role of availability in the estimation of national populations. *Memory & Cognition, 20,* 406–412. doi:10.3758/BF03210924

Brown, N. R., & Siegler, R. S. (1993). Metrics and mappings: A framework for understanding real-world quantitative estimation. *Psychological Review, 100,* 511–534. doi:10.1037/0033-295X.100.3.511

Brown, N. R., & Tan, S. (2011). Magnitude comparison revisited: An alternative approach to binary choice under uncertainty. *Psychonomic Bulletin & Review, 18,* 392–398. doi:10.3758/s13423-011-0057-1

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology, 103,* 1131–1136. doi:10.1037/h0037361

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100,* 432–459. doi:10.1037/0033-295X.100.3.432

Canadian Newsstand. (2009). [Database]. Retrieved from http://www.proquest.com/en-US/catalogs/databases/detail/canadian_newsstand.shtml

Cattell, J. McK. (1902). The time of perception as a measure of differences in intensity. *Philosophische Studien, 19,* 63–68.

Central Intelligence Agency. (2008a). Country comparison: GDP-per capita. In *The world factbook*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html

Central Intelligence Agency. (2008b). Country comparison: Population. In *The world factbook*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/rankorder/2119rank.html

Central Intelligence Agency. (2009). Country comparison: GDP-per capita. In *The world factbook*. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/rankorder/2004rank.html

Collins, A., & Michalski, R. (1989). The logic of plausible reasoning: A core theory. *Cognitive Science, 13,* 1–49. doi:10.1207/s15516709cog1301_1

Collins, A., Warnock, E. H., Aiello, N., & Miller, M. L. (1975). Reasoning from incomplete knowledge. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 383–415). New York, NY: Academic Press.

Dashiell, J. F. (1937). Affective value-distances as a determinant of esthetic judgment-times. *The American Journal of Psychology, 50,* 57–67. doi:10.2307/1416620

Dean, G. M., Dewhurst, S. A., Morris, P. E., & Whittaker, A. (2005). Selective interference with the use of visual images in the symbolic distance paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1043–1068. doi:10.1037/0278-7393.31.5.1043

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 626–641. doi:10.1037/0096-1523.16.3.626

Dougherty, M. R., Franco-Watkins, A. M., & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and fast and frugal heuristics. *Psychological Review, 115,* 199–211. doi:10.1037/0033-295X.115.1.199

Duncan, E. M., & McFarland, C. E., Jr. (1980). Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory & Cognition, 8,* .612–622. doi:10.3758/BF03213781

Erdfelder, E., Küpper-Tetzel, C. E., & Mattern, S. D. (2011). Threshold models of recognition and the recognition heuristic. *Judgment and Decision Making, 6,* 7–22. Retrieved from http://journal.sjdm.org/vol6.1.html

Friedman, A. (1978). Memorial comparisons without the "mind's eye". *Journal of Verbal Learning and Verbal Behavior, 17,* 427–444. doi:10.1016/S0022-5371(78)90256-6

Friedman, A., & Brown, N. R. (2000a). Reasoning about geography. *Journal of Experimental Psychology: General, 129,* 193–219. doi:10.1037/0096-3445.129.2.193

Friedman, A., & Brown, N. R. (2000b). Updating geographical knowledge: Principles of coherence and inertia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 900–914. doi:10.1037/0278-7393.26.4.900

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin, 113,* 44–66. doi:10.1037/0033-2909.113.1.44

Friedman, W. J., & deWinstanley, P. A. (2006). The mental representation of countries. *Memory, 14,* 853–871. doi:10.1080/09658210600782925

Fryman, J. F., & Wallace, J. (1985). Distorted cognitive maps: College students' misperceptions of nation size. *Perceptual and Motor Skills, 60,* 419–423. doi:10.2466/pms.1985.60.2.419

Gentner, D., & Collins, A. (1981). Studies of inference from lack of knowledge. *Memory & Cognition, 9,* 434–443. doi:10.3758/BF03197569

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science, 1,* 107–143. doi:10.1111/j.1756-8765.2008.01006.x

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62,* 451–482. doi:10.1146/annurev-psych-120709-145346

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103,* 650–669. doi:10.1037/0033-295X.103.4.650

Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making, 6,* 100–121. Retrieved from http://journal.sjdm.org/vol6.1.html

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98,* 506–528. doi:10.1037/0033-295X.98.4.506

Gigerenzer, G., Todd, P. M., & The ABC Research Group. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.

Glöckner, A., & Bröder, A. (2011). Processing of recognition information and additional cues: A model-based analysis of choice, confidence, and response time. *Judgment and Decision Making, 6,* 23–42. Retrieved from http://journal.sjdm.org/vol6.1.html

Goffin, R. D., & Olson, J. M. (2011). Is it all relative? Comparative judgments and the possible improvement of self-ratings and ratings of others. *Perspectives on Psychological Research, 6,* 48–60. doi:10.1177/1745691610393521

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109,* 75–90. doi:10.1037/0033-295X.109.1.75

Graesser, A. C., & Franklin, S. P. (1990). QUEST: A cognitive model of question answering. *Discourse Processes, 13,* 279–303. doi:10.1080/01638539009544760

Henmon, V. A. C. (1906). The time of perception as a measure of differences in sensations. *Archives of Philosophy, Psychology, and Scientific Methods, 8,* 1–75.

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1191–1206. doi:10.1037/a0013025

Hilbig, B. E. (2010). Reconsidering 'evidence' for fast and frugal heuristics. *Psychonomic Bulletin & Review, 17,* 923–930. doi:10.3758/PBR.17.6.923

Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2010). One-reason decision making unveiled: A measurement model of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 123–134. doi:10.1037/a0017518

Hilbig, B. E., & Pohl, R. F. (2009). Ignorance- versus evidence-based decision making: A decision time analysis of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1296–1305. doi:10.1037/a0016565

Hilbig, B. E., Pohl, R. F., & Bröder, A. (2009). Criterion knowledge: A moderator of using the recognition heuristic? *Journal of Behavioral Decision Making, 22,* 510–522. doi:10.1002/bdm.644

Holyoak, K. J. (1977). The form of analog size information in memory. *Cognitive Psychology, 9,* 31–51. doi:10.1016/0010-0285(77)90003-2

Holyoak, K. J., & Walker, J. H. (1976). Subjective magnitude information in semantic orderings. *Journal of Verbal Learning and Verbal Behavior, 15,* 287–299. doi:10.1016/0022-5371(76)90026-8

Honda, H., Abe, K., Matsuka, T., & Yamagishi, K. (2011). The role of familiarity in binary choice inference. *Memory & Cognition, 39,* 851–863. doi:10.3758/s13421-010-0057-9

Huntley-Fenner, G., & Cannon, E. (2000). Preschoolers' magnitude comparisons are mediated by a preverbal analog mechanism. *Psychological Science, 11,* 147–152. doi:10.1111/1467-9280.00230

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. doi:10.1016/j.jml.2007.11.007

Johnson, D. M. (1939). Confidence and speed in the two-category judgment. *Archives of Psychology, 241,* 1–52.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153. doi:10.1037/0033-295X.93.2.136

Kerst, S. M., & Howard, J. H., Jr. (1977). Mental comparisons for ordered information on abstract and concrete dimensions. *Memory & Cognition, 5,* 227–234. doi:10.3758/BF03197366

Kosslyn, S. M., Murphy, G. L., Bemesderfer, M. E., & Feinstein, K. J. (1977). Category and continuum in mental comparison. *Journal of Experimental Psychology: General, 106,* 341–375. doi:10.1037/0096-3445.106.4.341

Lee, M. D., & Cummins, T. D. R. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review, 11,* 343–352. doi:10.3758/BF03196581

Lee, P. J., & Brown, N. R. (2004). The role of guessing and boundaries on date estimation biases. *Psychonomic Bulletin & Review, 11,* 748–754. doi:10.3758/BF03196630

Leth-Steensen, C., & Marley, A. A. J. (2000). A model of response time effects in symbolic comparison. *Psychological Review, 107,* 62–100. doi:10.1037/0033-295X.107.1.162

Lichtenstein, S., & Slovic, P. (Eds.). (2006). *The construction of preference*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511618031

Lichtenstein, S., Slovic, P., Fischhoff, B., Layman, M., & Combs, B. (1978). Judged frequency of lethal events. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 551–578. doi:10.1037/0278-7393.4.6.551

Link, S. W. (1990). Modeling imageless thought: The relative judgment theory of numerical comparisons. *Journal of Mathematical Psychology, 34,* 2–41. doi:10.1016/0022-2496(90)90010-7

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika, 40,* 77–105. doi:10.1007/BF02291481

Maki, R. (1981). Categorization and distance effects with spatial linear orders. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 15–32. doi:10.1037/0278-7393.7.1.15

Maki, R. (1982). Why do categorization effects occur in comparative judgment tasks? *Memory & Cognition, 10,* 252–264. doi:10.3758/BF03197637

Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G., & Gigerenzer, G. (2010). From recognition to decisions: Extending and testing recognition-based models for multialternative inference. *Psychonomic Bulletin & Review, 17,* 287–309. doi:10.3758/PBR.17.3.287

Marewski, J. N., & Mehlhorn, K. (2011). Using the ACT-R architecture to specify 39 quantitative process models of decision making. *Judgment and Decision Making, 6,* 439–519. Retrieved from http://journal.sjdm.org/vol6.6.html

Marewski, J. N., Pohl, R. F., & Vitouch, O. (Eds.). (2010). Recognition processes in inferential decision making [Special issue]. *Judgment and Decision Making, 5*(4). Retrieved from http://journal.sjdm.org/vol5.4.html

Marewski, J. N., Pohl, R. F., & Vitouch, O. (Eds.). (2011a). Recognition processes in inferential decision making (II) [Special issue]. *Judgment and Decision Making, 6*(1). Retrieved from http://journal.sjdm.org/vol6.1.html

Marewski, J. N., Pohl, R. F., & Vitouch, O. (Eds.). (2011b). Recognition processes in inferential decision making (III) [Special issue]. *Judgment and Decision Making, 6*(5). Retrieved from http://journal.sjdm.org/vol6.5.html

Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review, 118,* 393–437. doi:10.1037/a0024143

Marshuetz, C. (2005). Order information in working memory: An integrative review of evidence from brain and behavior. *Psychological Bulletin, 131,* 323–339. doi:10.1037/0033-2909.131.3.323

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278. doi:10.1037/0033-295X.100.2.254

Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics, 13,* 180–184. doi:10.3758/BF03214124

Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology, 8,* 228–246. doi:10.1016/0010-0285(76)90025-6

Moyer, R. S., & Dumais, S. T. (1978). Mental comparison. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12, pp. 117–155). New York, NY: Academic Press.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215,* 1519–1520. doi:10.1038/2151519a0

Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review, 110,* 472–489. doi:10.1037/0033-295X.110.3.472

Mussweiler, T., & Epstude, K. (2009). Relatively fast! Efficiency advantages of comparative thinking. *Journal of Experimental Psychology: General, 138,* 1–21. doi:10.1037/a0014374

Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Sciences, 9,* 11–15. doi:10.1016/j.tics.2004.11.005

Newell, B. R., & Fernandez, D. (2006). On the binary quality of recognition and the inconsequentiality of further knowledge: Two critical tests of the recognition heuristic. *Journal of Behavioral Decision Making, 19,* 333–346. doi:10.1002/bdm.531

Newell, B. R., & Lee, M. D. (2011). The right tool for the job? Comparing an evidence accumulation and a naive strategy selection model of decision making. *Journal of Behavioral Decision Making, 24,* 456–481. doi:10.1002/bdm.703

Newell, B. R., & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 923–935. doi:10.1037/0278-7393.30.4.923

Nickerson, R. S. (1981). Motivated retrieval from archival memory. In H. E. Howe, Jr. & J. H. Flowers (Eds.), *Nebraska symposium on motivation: Cognitive processes* (Vol. 28, pp. 73–119). Lincoln: University of Nebraska Press.

Oppenheimer, D. M. (2003). Not so fast! (and not so frugal!): Rethinking the recognition heuristic. *Cognition, 90,* B1–B9. doi:10.1016/S0010-0277(03)00141-0

Pachur, T., Bröder, A., & Marewski, J. N. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making, 21,* 183–210. doi:10.1002/bdm.581

Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 983–1002. doi:10.1037/0278-7393.32.5.983

Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. G. (2011). The recognition heuristic: A review of theory and tests. *Frontiers in Cognitive Science, 2,* 147. doi:10.3389/fpsyg.2011.00147

Paivio, A. (1975). Perceptual comparisons through the mind's eye. *Memory & Cognition, 3,* 635–647. doi:10.3758/BF03198229

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139173933

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-Plus*. New York, NY: Springer. doi:10.1007/978-1-4419-0318-1

Pleskac, T. J. (2007). A signal detection analysis of the recognition heuristic. *Psychonomic Bulletin & Review, 14,* 379–391. doi:10.3758/BF03194081

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review, 117,* 864–901. doi:10.1037/a0019737

Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making, 19,* 251–271. doi:10.1002/bdm.522

Pohl, R. F. (2011). On the use of recognition in inferential decision making: An overview of the debate. *Judgment and Decision Making, 6,* 423–438. Retrieved from http://journal.sjdm.org/vol6.5.html

Poltrock, S. E. (1989). A random walk model of digit comparison. *Journal of Mathematical Psychology, 33,* 131–162. doi:10.1016/0022-2496(89)90027-8

Potts, G. R., Banks, W. P., Kosslyn, S. M., Moyer, R. S., Riley, C. A., & Smith, K. H. (1978). Encoding and retrieval in comparative judgments. In N. J. Castellan, Jr. & F. Restle (Eds.), *Cognitive theory* (Vol. 3, pp. 243–308). Hillsdale, NJ: Lawrence Erlbaum.

R Development Core Team. (2012). *R: A language and environment for statistical computing* (Version 2.15.1) [Software]. Vienna, Austria: R Foundation for Statistical Computing.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19,* 90–138. doi:10.1016/0010-0285(87)90005-3

Richter, T., & Späth, P. (2006). Recognition is used as one cue among others in judgment and decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 150–162. doi:10.1037/0278-7393.32.1.150

Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General, 135,* 207–236. doi:10.1037/0096-3445.135.2.207

Rilling, M., & McDiarmid, C. (1965). Signal detection in fixed-ratio schedules. *Science, 148,* 526–527. doi:10.1126/science.148.3669.526

Rosburg, T., Mecklinger, A., & Frings, C. (2011). When the brain decides: A familiarity-based approach to the recognition heuristic as evidenced by event-related brain potentials. *Psychological Science, 22,* 1527–1534. doi:10.1177/0956797611417454

Sailor, K. M., & Shoben, E. J. (1996). The role of categorical information in processing relational attributes. *Memory & Cognition, 24,* 756–765. doi:10.3758/BF03201100

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review, 112,* 610–628. doi:10.1037/0033-295X.112.3.610

Schweickart, O., Brown, N. R., & Lee, P. J. (2009, November). *On the role of recognition and magnitude-comparison in binary decision tasks*. Paper presented at the annual meeting of the Society for Judgment and Decision Making, Boston, MA.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development, 48,* 630–633. doi:10.2307/1128664

Siegler, R. S., & Robinson, M. (1982). The development of numerical understandings. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior* (Vol. 16, pp. 242–312). New York, NY: Academic Press.

Temple, E., & Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proceedings of the National Academy of Sciences, 95,* 7836–7841. doi:10.1073/pnas.95.13.7836

Todd, P. M., Gigerenzer, G., & The ABC Research Group. (2012). *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195315448.001. 0001

Vickers, D. (1979). *Decision processes in visual perception*. New York, NY: Academic Press.

Volz, K. G., Schooler, L. J., Schubotz, R. I., Raab, M., Gigerenzer, G., & von Cramon, D. Y. (2006). Why you think Milan is larger than Modena: Neural correlates of the recognition heuristic. *Journal of Cognitive Neuroscience, 18,* 1924–1936. doi:10.1162/jocn.2006.18.11.1924

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74* B1–B11. doi:10.1016/S0010-0277(99)00066-9

Appendix A

Countries Used in Study 2

Set U1	Set U2	Set R1	Set R2
Benin	Eritrea	Nepal	Brazil
Burkina Faso	The Gambia	Ethiopia	Cambodia
Mauritania	Mauritius	Bulgaria	India
Moldova	Namibia	Kazakhstan	Philippines
Oman	Tajikistan	Egypt	Ukraine
Azerbaijan	Ivory Coast	Haiti	Uganda
Kyrgyzstan	Gabon	Pakistan	Nicaragua
Lesotho	Burundi	China	Syria

Appendix B

Countries Used in Study 3 (Fixed Set)

Likely unknown countries	Likely known countries
Benin	Brazil
Eritrea	Cambodia
Kyrgyzstan	Ukraine
Mauritania	Ethiopia
Oman	Philippines
Burkina Faso	Egypt
Azerbaijan	Uganda
Lesotho	Kazakhstan
Mauritius	Haiti
Moldova	India

Received May 2, 2012
Revision received October 4, 2012
Accepted November 26, 2012 ■