Cognition 157 (2016) 77–99

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Strategies for memory-based decision making: Modeling behavioral and neural signatures within a cognitive architecture

Hanna B. Fechner [a,b,*], Thorsten Pachur [c], Lael J. Schooler [d,a], Katja Mehlhorn [e,b], Ceren Battal [f,g], Kirsten G. Volz [g], Jelmer P. Borst [e,b]

[a] Max Planck Institute for Human Development, Center for Adaptive Behavior and Cognition, Lentzeallee 94, 14195 Berlin, Germany
[b] Carnegie Mellon University, Department of Psychology, 5000 Forbes Ave, Pittsburgh, PA 15213, USA
[c] Max Planck Institute for Human Development, Center for Adaptive Rationality, Lentzeallee 94, 14195 Berlin, Germany
[d] Syracuse University, Department of Psychology, 301 Hall of Languages, Syracuse, NY 13244, USA
[e] University of Groningen, Department of Artificial Intelligence, Nijenborgh 9, 9747 AG Groningen, The Netherlands
[f] University of Trento, Center for Mind/Brain Sciences, Corso Bettini 31, 38068 Rovereto, TN, Italy
[g] University of Tübingen, Werner Reichardt Centre for Integrative Neuroscience, Otfried-Müller-Straβe 25, 72076 Tübingen, Germany

ARTICLE INFO

Article history:
Received 23 July 2015
Revised 15 August 2016
Accepted 17 August 2016

Keywords:
Decision making
Memory
Recognition heuristic
ACT-R
Computational modeling
Neuroimaging

ABSTRACT

How do people use memories to make inferences about real-world objects? We tested three strategies based on predicted patterns of response times and blood-oxygen-level-dependent (BOLD) responses: one strategy that relies solely on recognition memory, a second that retrieves additional knowledge, and a third, lexicographic (i.e., sequential) strategy, that considers knowledge conditionally on the evidence obtained from recognition memory. We implemented the strategies as computational models within the Adaptive Control of Thought-Rational (ACT-R) cognitive architecture, which allowed us to derive behavioral and neural predictions that we then compared to the results of a functional magnetic resonance imaging (fMRI) study in which participants inferred which of two cities is larger. Overall, versions of the lexicographic strategy, according to which knowledge about many but not all alternatives is searched, provided the best account of the joint patterns of response times and BOLD responses. These results provide insights into the interplay between recognition and additional knowledge in memory, hinting at an adaptive use of these two sources of information in decision making. The results highlight the usefulness of implementing models of decision making within a cognitive architecture to derive predictions on the behavioral and neural level.

1. Introduction

When judging which alternative scores higher on some feature (i.e., the criterion of interest), people often search for useful information (i.e., cues) in their memories (e.g., Gigerenzer, Hoffrage, & Kleinbölting, 1991). For instance, how do people choose chocolate candies from a wide selection as a gift for a friend? To choose the ones the friend will enjoy most (criterion of interest) people might select candies based solely on whether they recognize their brand name, or they might delve deeper into their memories and retrieve additional information (e.g., how much they themselves enjoyed the candies, whether the company was involved in a recent scandal, etc.).

There has been considerable debate as to which strategies people use when making such decisions from memory (e.g., Hilbig, Erdfelder, & Pohl, 2011; Marewski, Gaissmaier, Schooler, Goldstein, & Gigerenzer, 2010; Newell & Shanks, 2004; Oppenheimer, 2003; Pachur, Bröder, & Marewski, 2008; Schwikert & Curran, 2014). One prominent proposal is that the recognition process for decision alternatives provides an important source of information (as illustrated in the example above). We refer to strategies that solely use recognition to make decisions as *recognition-based strategies* (e.g., Goldstein & Gigerenzer, 2002; Hertwig, Herzog, Schooler, & Reimer, 2008; Schooler & Hertwig, 2005; for other interpretations see, e.g., Tversky & Kahneman, 1973; Whittlesea, 1993). Another potential source of information is knowledge about the decision alternatives stored in long-term memory. We refer to strategies that search for knowledge beyond

* Corresponding author at: Max Planck Institute for Human Development, Center for Adaptive Behavior and Cognition, Lentzeallee 94, 14195 Berlin, Germany.

E-mail addresses: hfechner@mpib-berlin.mpg.de (H.B. Fechner), pachur@mpib-berlin.mpg.de (T. Pachur), lschoole@syr.edu (L.J. Schooler), s.k.mehlhorn@rug.nl (K. Mehlhorn), ceren.battal@unitn.it (C. Battal), kirsten.volz@cin.uni-tuebingen.de (K.G. Volz), j.p.borst@rug.nl (J.P. Borst).

http://dx.doi.org/10.1016/j.cognition.2016.08.011
0010-0277/

recognition as *knowledge-based strategies* (e.g., Gigerenzer & Goldstein, 1996). Finally, people might use both recognition and knowledge as cues, such that they first consider recognition, and if that does not seem to provide a reliable basis for the decision, they move on to knowledge (cf. Erdfelder, Küpper-Tetzel, & Mattern, 2011; Marewski et al., 2010; Schwikert & Curran, 2014).

Our goal in this article was to test these different decision mechanisms by implementing them as computational models within one common framework—the ACT-R (Adaptive Control of Thought–Rational) cognitive architecture (Anderson, 2007)—and comparing their outcomes to empirical response times and functional magnetic resonance imaging (fMRI) data. We acknowledge that there are alternative frameworks for modeling decisions and response times (e.g., evidence accumulation: Lee & Cummins, 2004; Pleskac & Busemeyer, 2010; and connectionist approaches: Glöckner, Hilbig, & Jeckel, 2014), but rely on ACT-R here because it also allows one to derive specific predictions for fMRI activation.

To achieve this goal, we collected behavioral and fMRI data in an inference task in which people judged which of two cities was larger. The city domain is convenient because people are likely to have naturally acquired recognition and knowledge about cities, and both types of information can be good indicators of the decision criterion (e.g., Pachur, Todd, Gigerenzer, Schooler, & Goldstein, 2011); the domain therefore provides a good test bed for investigating how they are used to make inferences in a real-world domain. In addition, using this domain allows for comparability with the many previous studies on memory-based decisions that have also used the city-size task (e.g., Goldstein & Gigerenzer, 2002; Hilbig et al., 2011; Horn, Ruggeri, & Pachur, in press; Marewski & Schooler, 2011; Pachur, Mata, & Schooler, 2009; Rosburg, Mecklinger, & Frings, 2011; Schwikert & Curran, 2014; Volz et al., 2006). We implemented the decision strategies as computational models within the ACT-R cognitive architecture (Anderson, 2007), which yields predictions at both the behavioral and the neural level for each of the candidate strategies.[1]

1.1. Recognition-based strategies

Recognition-based strategies exploit the process of recognizing the decision alternatives to make an inference. Two well-studied instances of recognition-based strategies are applicable in different situations. The first is the *recognition heuristic* (Goldstein & Gigerenzer, 1999, 2002; for reviews see Gigerenzer & Goldstein, 2011; Pachur et al., 2011).[2] The recognition heuristic is assumed to be an adaptive cognitive tool that is applied when there is a (strong) positive relationship between recognition and the criterion (e.g., the number of inhabitants of a city in the present study) in the environment. It then predicts that if one of two alternatives (i.e., cities in the present study) is recognized, but not the other, then it is inferred that the recognized alternative has a higher value on the criterion.[3] According to the recognition heuristic, recognition is used in a noncompensatory way, which means that recognition cannot be overruled by any other, possibly contradictory cue knowledge.

Second, the *fluency heuristic* (e.g., Hertwig et al., 2008; Schooler & Hertwig, 2005) predicts that if both alternatives (i.e., cities) are recognized, but one was recognized faster (i.e., more fluently), then it is inferred that the more fluently recognized alternative has a higher value on the criterion. The fluency heuristic exploits the perceived times for successful retrievals of city names from long-term memory. Previous research has shown that exploiting recognition and recognition times can lead to accurate inferences in many real-world domains (e.g., geography, sports, politics, economics; Herzog & Hertwig, 2013; Pachur et al., 2011), among them the city domain (Goldstein & Gigerenzer, 1999). Based on these accounts, we tested a recognition-based strategy that attempts to recognize both alternatives. When only one of the alternatives is recognized, it chooses the recognized one and when both alternatives are recognized, it chooses the one that is recognized faster.

1.2. Knowledge-based strategies

After attempting to recognize the alternatives, knowledge-based strategies retrieve additional knowledge to make a decision. Such strategies can rely on the most important piece of knowledge (e.g., Gigerenzer & Goldstein, 1996) or integrate the available and relevant pieces (e.g., Dawes, 1979; Payne, Bettman, & Johnson, 1993). In the city example, people might retrieve the information that a city has an international airport, significant industry, or a university (cf. Pachur et al., 2008). In the current study we made no specific assumptions about the algorithm by which such individual pieces of knowledge are processed or integrated but instead focused on the general question of whether any additional knowledge beyond recognition was retrieved before the decision. Therefore, we tested a strategy that searches for knowledge about each recognized alternative before making a decision.

1.3. A strategy for the lexicographic use of recognition and knowledge

A third strategy we tested allows for a lexicographic (i.e., sequential) consideration of recognition (R) and additional knowledge (K). We therefore refer to this strategy as *Lex-R-K*. Some studies have obtained only limited support for a strict use of the recognition and fluency heuristics (e.g., Hilbig, Erdfelder, & Pohl, 2010; Hilbig et al., 2011; Schwikert & Curran, 2014) and purely knowledge-based strategies (e.g., Marewski et al., 2010; Pachur & Biele, 2007), so strategies implementing a confluence of different types of memory information might be more appropriate. Lex-R-K draws on proposals that the recognition cue might be evaluated in terms of memory strength (indicating certainty of recognition memory) before it is used as a basis for a decision (Erdfelder et al., 2011; Marewski et al., 2010; Schwikert & Curran, 2014).

We extended these ideas into a mechanism that evaluates whether the recognition information—for one as well as two recognized alternatives—is sufficiently reliable to be used as a basis for a decision; if not, it moves on to additional knowledge (for alternative proposals see Marewski & Mehlhorn, 2011). As a proxy for the reliability of recognition information, the strategy evaluates the speed with which decision alternatives are recognized, which is an indicator of memory strength (for evidence, see, e.g., Pleskac & Busemeyer, 2010; Van Zandt, 2000).

When only one alternative (i.e., city) is recognized, Lex-R-K checks if this alternative was recognized sufficiently quickly—that is, if the recognition time is equal to or falls below a certain threshold (see also Erdfelder et al., 2011)—indicating that recognition provides a reliable cue for the decision. When both alternatives are recognized, the strategy evaluates if one was recognized

[1] We use the term "predictions" interchangeably with "simulation results" of models, as is common in the literature. Nevertheless, note that our models were created after collecting the data and use parts of these data to account for the contents and retrieval dynamics of participants' memories.

[2] The recognition heuristic assumes a binary recognition state. There is some evidence that the distinction between recognized and not recognized might be too simplistic (e.g., Bröder & Schütz, 2009) and that an additional uncertainty state is necessary to account for the data. Erdfelder et al. (2011) proposed an extension of the recognition heuristic that assumes such a tertiary structure of recognition memory.

[3] When there is no or only a weak correlation, the heuristic is assumed to be used less or not at all, and several studies have found support for this assumption (Pachur et al., 2009, 2011; Pohl, 2006). When there is a negative correlation between recognition and the criterion (which is the case, for instance, when one reverses the question from "which is the larger city?" to "which is the smaller city?") it is thus predicted that people choose the unrecognized object. This also seems to correspond to how people use recognition (e.g., Frosch, McCloy, Beaman, & Goddard, 2015; McCloy, Beaman, Frosch, & Goddard, 2010).

sufficiently faster than the other—that is, if the difference in their recognition time exceeds a particular threshold (cf. Luan, Schooler, & Gigerenzer, 2014). If one of these cases applies, then recognition information is used to make a decision. Otherwise, a search for additional knowledge is initiated. The existence or nonexistence of retrievable knowledge about an alternative can then serve as a basis for the decision. Thus, Lex-R-K depends on the setting of two thresholds: if only one alternative was recognized, an *absolute threshold* to determine whether the recognition time was sufficiently fast, and if both alternatives were recognized, a *difference threshold* to determine whether the recognition time difference was sufficiently large. We tested three versions of Lex-R-K, which assume different values for these thresholds such that Lex-R-K searches for additional knowledge in a small, intermediate, or large number of cases.

1.4. Using a cognitive architecture to derive predictions about behavioral and neural data

To test the strategies, we implemented them as computational models within the ACT-R cognitive architecture (Anderson, 2007), which allowed us to derive predictions for response times and blood-oxygen-level-dependent (BOLD) activation differences between trial types (e.g., whether both, one, or neither city was recognized, and whether additional cue knowledge was available). ACT-R is an integrated theory of cognition and a modeling environment that has been applied to a wide range of human behavior such as perception, learning and memory, language processing, problem solving, decision making, cognitive development, and the design of intelligent tutoring systems (for a list of publications see http://act-r.psy.cmu.edu/publication/). In this architecture each ACT-R model consists of a set of production rules and declarative knowledge that interact with core modules representing cognitive functions, such as vision, long-term memory retrieval, working memory updating, and motor responses.

The activity of ACT-R's core modules has been mapped to different regions of interest (ROIs) in the brain (Anderson, 2007; Anderson, Fincham, Qin, & Stocco, 2008; for a meta-analysis of five different tasks and models, see Borst, Nijboer, Taatgen, van Rijn, & Anderson, 2015). It has been proposed that when a module of the architecture is active, the corresponding brain region shows an increase in BOLD signal. The claim is not that the ROIs are the only parts of the brain that are activated when the modules are operating, but that these regions are good indicators of the modules' activity. The increase in BOLD activation in an ROI can be calculated by convolving the demand function that indicates when a module is active with a hemodynamic response function (see for details, e.g., Borst & Anderson, 2015). Thus, the predefined mapping of ACT-R modules onto brain regions allows researchers to derive quantitative predictions for the BOLD response in these ROIs, and consequently to constrain models based on fMRI experiments—as we did in the current study.

The models of the three strategies result in different BOLD predictions for two ACT-R modules. First, the activity of the retrieval module reflects the retrieval of information from declarative memory. It was mapped to a region around the inferior frontal gyrus in the prefrontal cortex (PFC; left panel of Fig. 1). Previous studies have found that activation of this region reflects both the amount of retrieved information and the difficulty of retrieving episodic memories (e.g., Anderson, Byrne, Fincham, & Gunn, 2008; Borst & Anderson, 2013; Danker, Gunn, & Anderson, 2008; Sohn, Goode, Stenger, Carter, & Anderson, 2003; Sohn et al., 2005). Such factors also impact the retrieval of real-world memories for cities and knowledge facts in our study. Second, activity of a module critical to the functioning of working memory (i.e., ACT-R's imaginal module, which maintains the problem state of task-relevant information) was mapped to a region close to the intraparietal sulcus in the posterior parietal cortex (PPC; right panel of Fig. 1). This region has been shown to be sensitive to updates to problem representations in working memory (e.g., Anderson, Albert, & Fincham, 2005; Anderson, Byrne, et al., 2008; Borst & Anderson, 2013; Borst, Taatgen, Stocco, & van Rijn, 2010; Borst, Taatgen, & van Rijn, 2011; Danker et al., 2008; Sohn et al., 2005). In our study, the problem representation contained information about the decision alternatives. This information was updated by retrieving information from declarative memory. For the purpose of our analysis, we refer to these brain areas as the *retrieval ROI* and the *working memory ROI*.

Using ACT-R models of the recognition-based strategy, the knowledge-based strategy, and three versions of Lex-R-K that search for knowledge in a small, intermediate, or large number of cases, we derived quantitative predictions for the behavioral and neural data that we collected. Qualitatively speaking, compared to the recognition-based strategy, the knowledge-based strategy involves (a) more memory retrievals, as it attempts to retrieve knowledge beyond recognition (which increases response time and activity in the retrieval ROI in the PFC); (b) more working memory updates, as it needs to store the retrieved information to have it available for the decision (which increases response time and activity in the working memory ROI in the PPC); and (c) more time to control and coordinate these actions (increasing response time). Depending on whether or not Lex-R-K moves on to consider additional knowledge in a given trial, it makes the same predictions as the knowledge-based or the recognition-based strategy, respectively. Therefore, on the aggregate level predictions of the versions of Lex-R-K fall between those of the recognition-based and knowledge-based models; where exactly they fall depends on the amount of knowledge that is searched and retrieved.

We used empirical response times and fMRI data to evaluate our computational models based on their differential predictions (for more information on this approach see, e.g., Borst et al., 2015; Turner, Forstmann, Love, Palmeri, & van Maanen, in press). Beyond the response time and BOLD predictions, all models in principle could also be used to derive predictions for choices. However, in the empirical study no information was collected about the cue knowledge that participants might have used for each specific decision, which would have been necessary to predict which alternative is chosen based on knowledge. We nevertheless analyzed the choice data, making the simplifying assumption that models that base their decision on knowledge would chose the alternative for which knowledge was available or—if there was knowledge about both alternatives—would choose the alternative for which the knowledge was retrieved faster.[4] As this assumption may often be incorrect, in our primary analyses we focused on response times and BOLD responses and used the choice data only to supplement these analyses.

We next report the empirical study in which participants performed the inference task that asked which of two cities had more inhabitants (the main task) and then a metacognitive judgment task (which is reported separately; Battal, Fechner, Schooler, & Volz, 2012) while positioned in an MRI scanner. Participants then worked outside the scanner on tasks that assessed their recognition of cities and the availability of knowledge about these cities. After reporting the empirical study, we describe the ACT-R models of the three tested strategies in detail and show how we derived predictions for response times and fMRI data. Finally, we compare

[4] This assumption relates to proposals that people may mainly encode and generate cue knowledge that provides positive evidence for a high value on the decision criterion (cf. Dougherty, Franco-Watkins, & Thomas, 2008) and the finding that such cue knowledge may be retrieved faster than cue knowledge that provides negative evidence (Marewski & Mehlhorn, 2011).

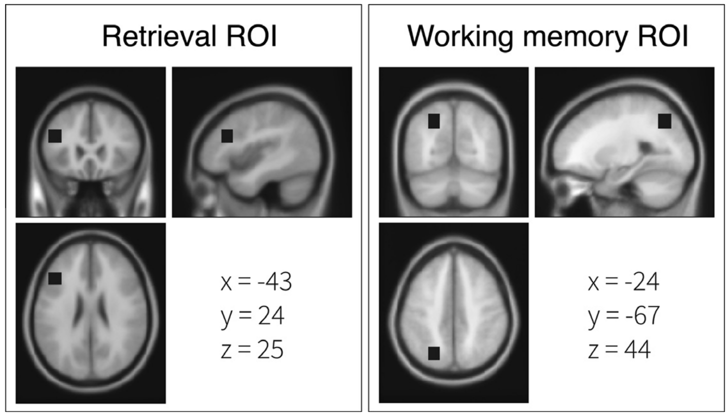

Fig. 1. The retrieval region of interest (ROI) and working memory ROI with their Montreal Neurological Institute (MNI) coordinates. The activity of ACT-R's module responsible for retrieval from declarative memory has been associated with the region in the left lateral inferior prefrontal cortex (left panel), and activity of the module responsible for working memory updating with the region in the left posterior parietal cortex (right panel).

and evaluate how well these predictions accounted for the observed quantitative and qualitative data patterns of the empirical study.

2. Method

2.1. Participants

Twenty-seven healthy, right-handed volunteers participated in the study (15 women, 12 men; M_{age} = 26.43 years, *SD* = 3.98; range 21–37 years). They were recruited via university-wide e-mails at the University of Tübingen. Informed consent was obtained according to the Declaration of Helsinki. The local ethics committee of the University of Tübingen approved the experimental procedure. Participants were paid 12 euros per hour. The duration of the fMRI experiment was about 1 h, and the subsequent behavioral part took 1.5 h, on average. Six of the original 33 participants were excluded: four because of technical problems during data collection (i.e., two during the functional scan, one during the anatomical scan, and one during the behavioral data collection), one because of a missing randomization in the trials, and one because, after preprocessing, the visual regions lay outside the field of view.

2.2. Materials

The inference task consisted of 170 trials: For 140 trials, cities were randomly sampled for each participant from a pool of 309 cities from seven countries; the pool of cities consisted of 20 from Canada, 35 from France, 48 from Great Britain, 29 from Italy, 20 from The Netherlands, 39 from Spain, and 118 from the United States. Only cities within each country were paired. In addition, 15 trials with well-known German cities were presented to include trials where both cities would be surely recognized. Each city appeared only once across trials—with the exception of Birmingham and Richmond, which were repeated once (but not within the same city pair). Fifteen null events were included where no cities but only a fixation cross was presented to let the BOLD response return to baseline.

2.3. Procedure

The first part of the study took place in the scanner and started with a preparation phase of 10,000 ms informing the participant that the experiment was about to begin. Then the 170 trials were presented in random order. Each trial lasted 10,000 ms in total, which corresponds to five scans with a repetition time (TR) of 2,000 ms. Each trial consisted of the inference task and the metacognitive judgment task and started with a fixation cross with a duration of either 10, 500, or 1,000 ms (randomly determined). The onset of the fixation cross was synchronized to the onset of the scanner pulse. Then the names of two cities were presented on the screen (Fig. 2). Cities with more inhabitants were allocated to the left and right position on the screen equally often. Participants saw the screen through a mirror system and were instructed to indicate which city had more inhabitants. They responded with the right index finger for the left city and with the right middle finger for the right city. Responding was possible for a maximum of 4,900 ms after the trial had started. After a response was given, the names of the cities were replaced by a fixation cross. This cross was shown until 4,900 ms of total trial duration had elapsed.

After each city pair, participants were presented with the metacognitive judgment task in which they were asked to indicate the basis for their decision (Fig. 2).[5] Responding was possible for a maximum of 5,100 ms after onset of the metacognitive judgment task. After the final response the contents of the screen were replaced by a fixation cross. This cross was shown until the full 10,000 ms of the trial had elapsed and the next trial started. In this article we focus on the results for the inference task.

In the second part of the study, participants worked on several tasks outside the scanner: In a recognition task, the names of all previously shown cities were presented again in random order and separated by fixation crosses of 1,000-ms duration. Participants indicated whether they recognized each city, that is, whether they had seen or heard the city name before the experiment. They responded by pressing keys on a standard German keyboard, using the right Alt key for recognized and the left Alt key for not recognized. The name of each city was visible on the screen until a

[5] In this task, participants had three response options: knowledge (K in the display, left ring finger), feeling of knowing (F, left middle finger), or guessing (G, left index finger). Participants could combine the response options in the order that best described their decision and indicated that they had finished with their left thumb (i.e., they could give multiple responses). The response options were explained to the participants as follows: The decision was based on knowledge if the number of inhabitants was known or further information was available on one or both cities from which they could infer which city had more inhabitants. The decision was based on feeling of knowing if they did not use knowledge for the decision, that is, if they could not give factual reasons why one city should have more inhabitants, but they had the feeling of knowing which city had more inhabitants. The decision was based on a guess if they had no knowledge or feeling about which city had more inhabitants and had to guess.

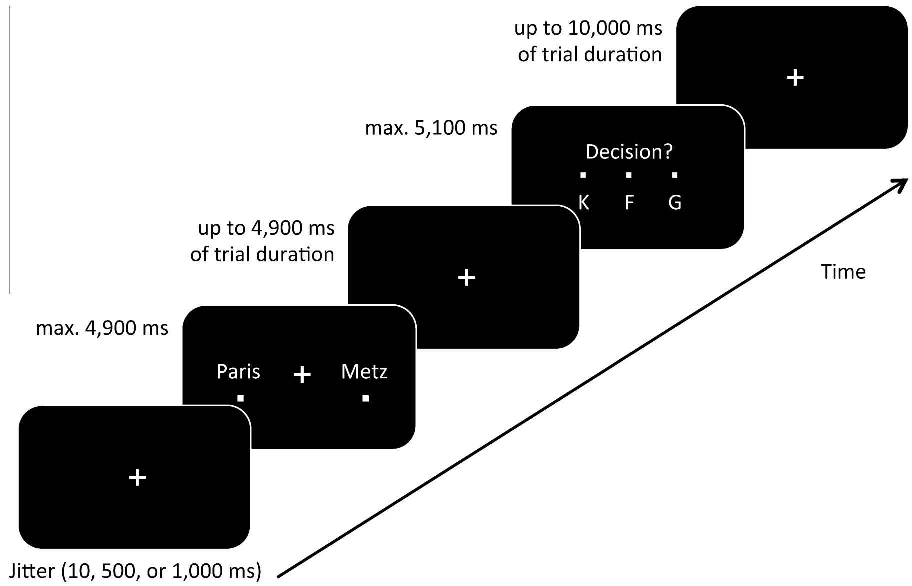

Fig. 2. Tasks in the scanner: Inference task (Paris + Metz) and metacognitive judgment task (Decision? K F G). In the inference task, participants had to decide which of two cities had more inhabitants. In the metacognitive judgment task, participants had to indicate for each decision of the inference task, whether their response was based on knowledge (K), feeling of knowing (F), or guessing (G).

response was given. In an additional knowledge task, the names of all cities that the participant indicated as recognized were presented again in random order and separated by fixation crosses of 1,000-ms duration. Participants indicated whether or not they had any kind of additional knowledge about each city by pressing the right or left Alt key, respectively. The name of each city was visible on the screen until a response was given.

For the entire study participants were instructed to respond quickly without committing avoidable errors. After each response, squares on the screen near the response options changed their color from white to blue for 400 ms to indicate to participants that their responses were recorded. A fixation cross was visible at the center of the screen throughout all tasks, except for the metacognitive judgment task (where the screen center was needed to display all response options). Words and the fixation cross were presented in white on a black background. For stimulus presentation and response recording in all tasks, the Presentation software (version 14.8 Build 12.30.10, Neurobehavioral Systems, Inc.) was used.

2.4. Acquisition of fMRI data

The functional images were acquired on a 3.0 T Siemens Magnetom Tim Trio scanner (Erlangen, Germany) by using a single-shot echo-planar-imaging sequence (2,000-ms TR, 30-ms echo time, TE, 90° flip angle, 210-mm field of view, 64×64 matrix size, 3×3-mm in-plane resolution). Each functional volume consisted of 30 axial slices (3-mm thickness, 0.5-mm gap between slices), covering the whole brain. One functional run with 860 time points was run with each time point sampling over the 30 slices, interleaved from bottom to top. After the functional scans an anatomical run with 192 time points was run (2,300-ms TR, 2.98-ms TE, 9° flip angle, 256-mm field of view, 256×256 data matrix, $1 \times 1 \times 1$-mm voxel size).

2.5. Analysis of fMRI data

The data were preprocessed using SPM 8 (Statistical Parametric Mapping, Wellcome Trust Centre for Neuroimaging, London) for MATLAB (version R2012a, TheMathWorks, Inc.). This included realignment of functional images, slice-time correction in reference to the middle image, coregistration with the anatomical images, segmentation and normalization with reference to the International Consortium for Brain Mapping (ICBM) space template for European brains, and smoothing with an $8 \times 8 \times 8$-mm full-width half-maximum Gaussian kernel. No subjects were excluded for head movements, as all recorded movements were smaller than 4 mm in all directions.

ROI analyses were conducted using the MarsBaR region of interest toolbox for SPM (Brett, Anton, Valabregue, & Poline, 2002). The focus was on two regions—depicted in Fig. 1—that have previously been linked to ACT-R's module activity: For the retrieval module we used a region in the left lateral inferior PFC (Montreal Neurological Institute, MNI, coordinates: $x = -43$, $y = 24$, $z = 25$; size: $5 \times 5 \times 5$ voxels). For the module responsible for working memory updating we used a region in the left PPC (MNI coordinates: $x = -24$, $y = -67$, $z = 44$; size: $5 \times 5 \times 6$ voxels).[6] Left regions were used as they represent module activity better than right regions, as shown by a meta-analysis of five different studies, ranging from associative recognition to high-level multitasking (Borst & Anderson, 2013; Borst et al., 2015). For all ROI results the first scan of each trial constituted the baseline from which percentage BOLD change is reported. Given that the duration of the fixation cross in the beginning of each trial differed (i.e., 10, 500, or 1,000 ms), and the onset of the fixation cross was synchronized to the scanner pulse, trials with different durations of the fixation cross differed in terms of their baseline and time course of raw BOLD values across the trial. To calculate an estimate of the BOLD signal starting from stimulus onset (i.e., the city names) independent of the duration of the

[6] The ROIs were developed for MarsBaR by Borst et al. (2010). There the authors used 3.125×3.125 voxels, with 3.2-mm thickness—leading to the millimeter size of $15.625 \times 15.625 \times 16$ mm. This in turn was based on previous ACT-R fMRI studies (e.g., Stocco & Anderson, 2008) that used the same voxels. On the basis of this previous work, for the present study we defined box-shaped ROIs in millimeter space, that is, $15.625 \times 15.625 \times 16$ mm. Applied to $3 \times 3 \times 3$-mm voxels of the MNI brain, this leads to $5 \times 5 \times 5$ voxels ($15 \times 15 \times 15$ mm) being picked for the retrieval ROI, and $5 \times 5 \times 6$ voxels ($15 \times 15 \times 18$ mm) for the working memory ROI; the sizes are different because of where the box is placed exactly with respect to the voxels.

fixation cross, we computed a weighted average of pairs of subsequent scans according to the duration of the respective fixation cross.[7] For data analysis we summed the percentage BOLD change across the five scans of each trial and averaged over the trials of each subject within each condition.

3. Computational models

3.1. Computational models of the decision strategies in ACT-R

All decision strategies were implemented as computational models in the ACT-R modeling framework (ACT-R 6.0, version 1.5 [r1577]), which consist of chains of production rules that reside in procedural memory (for other examples see, e.g., Marewski & Mehlhorn, 2011; see also Marewski & Schooler, 2011; Schooler & Hertwig, 2005). To simulate cognition, these production rules interact with different cognitive resources (i.e., ACT-R's modules), such as for vision, retrieval from declarative memory, working memory updating, or motor responses. Modules have buffers that store information in the form of chunks that represent small units of information. The production rules check this information before initiating actions. These actions change the contents of the buffers and thus chains of behavior (e.g., a particular decision strategy) can be implemented. Here, we describe the models of the strategies tested by focusing on their main processing steps and especially those steps that are responsible for their predictions. Fine-grained flowcharts of the strategies, illustrating their processing steps as implemented by the different ACT-R modules, are shown in Appendix A; details on computations and parameter settings of ACT-R's modules (which were constant across all models) are presented in Appendix B.

In the model of the recognition-based strategy, at the start of each trial visual attention is shifted to each of the two city names, and an attempt is made to recognize each city by retrieving it from memory; the recognition times are stored in working memory and are used to make the decision. The knowledge-based model extends the memory search: After storing the recognition times for the cities, it shifts visual attention again to the name of each recognized city, tries to retrieve knowledge about these cities using the city names as retrieval cues, and stores the retrieval times for knowledge in working memory. Then the model makes a decision.

The Lex-R-K model implements a combination of the recognition-based and the knowledge-based model. It stores the recognition times of cities and evaluates them: When one city was recognized, the model checks how quickly this occurred. If recognition time was equal to or below the absolute threshold, Lex-R-K makes a decision, just as in the recognition-based model. If recognition time was above the absolute threshold, Lex-R-K attempts to retrieve knowledge for the recognized city, just like the knowledge-based model. When both cities were recognized, Lex-R-K evaluates if there was a sufficiently large difference in recognition time (i.e., above the difference threshold). If so, the model makes a decision. If the difference was too small (i.e., equal to or below the difference threshold), the model searches for knowledge about both cities and then makes a decision. All models incorporate guessing mechanisms to handle cases where decisions could not be made on the basis of memory.[8]

All models were run 10 times (to reduce random fluctuations in the predictions in those cases in which the models predicted a guessing response) on the trials of the inference task that each participant had seen and evaluated based on a comparison of the model predictions to the observed data in terms of response times and the BOLD responses. As indices of model performance, we used the root-mean-square deviation (*RMSD*) and the correlation coefficient *r*, which quantify the quantitative and qualitative correspondence, respectively, between the predicted and observed patterns.[9] Note that for none of the models parameters were adjusted to fit the data in the inference task; therefore, issues of model generalizability (that might arise due to overfitting) do not apply (e.g., Pitt, Myung, & Zhang, 2002). The models take into account that the response times and BOLD responses are impacted not only by the quantity of retrieved memories, but also by the accessibility of each individual memory. This memory information, as well as three a priori defined threshold levels for the three versions of Lex-R-K (which made Lex-R-K search for additional knowledge in a small, intermediate, or large number of cases), was determined based on participants' responses in the separate data sets of the recognition task and additional knowledge task (for details, see Section 3.2 and Appendix B).

In the model simulations, it was assumed that after each decision in the inference task a response was given in the metacognitive judgment task. However, this did not affect the model predictions of response times in the inference task, as the metacognitive task was completed after the inference task. It also did not affect the predicted BOLD patterns for the retrieval ROI in the PFC and the working memory ROI in the PPC because for the metacognitive judgment task, the models do not rely on memory retrieval and working memory updating.

3.2. Dynamics of memory retrieval in ACT-R

A challenge for the models is that we applied them to acccount for decisions that are based on real-world memories, formed outside of the laboratory. Real-world memories differ in terms of their accessibility and retrieval time. In general, ACT-R models naturally accommodate variability in the retrieval time for individual memories: In ACT-R the speed of retrieving a memory is a function of its activation. This activation is driven by the frequency and recency of past encounters with the contents of this memory in the world (Anderson, Fincham, & Douglass, 1999; Anderson & Schooler, 1991). For the city example, memories for often-encountered cities and knowledge about those cities have a higher activation and faster retrieval time compared to memories with fewer or less recent encounters.

To take these dynamics of memory retrieval into account, we measured the real-world memories for cities and knowledge for each participant in the recognition and knowledge task and used them as the basis for populating ACT-R's declarative memory. To do so, we estimated the retrieval times for the memories (i.e., the recognition and rejection times for city names as well as the retrieval and retrieval failure times for knowledge) by subtracting estimated nonretrieval time (consisting of processes for visual attention, action coordination, and the motor response) from the response times in the recognition and knowledge tasks and used

[7] In detail, we used a linear interpolation between measurements obtained from subsequent scans; that is, for the first interpolated scan we computed the BOLD response at Scan 1 $*$ (1 − the duration of the jittered fixation cross/TR) + the BOLD response at Scan 2 $*$ (the duration of the jittered fixation cross/TR), and then for the second interpolated scan we computed the BOLD response at Scan 2 $*$ (1 − the duration of the jittered fixation cross/TR) + the BOLD response at Scan 3 $*$ (the duration of the jittered fixation cross/TR), and so forth. The results did not change in any qualitative way without this correction.

[8] For guessing, each model has two production rules that can fire when the conditions for guessing are met. One production rule chooses the first and the other the second city. Which production rule fires was determined randomly.

[9] We relied on these indices of model performance rather than likelihood measures as it is not straightforward to derive likelihoods for the models in the present case; this is because the models are deterministic in their predictions for response times and BOLD responses on each trial.

these estimated times in the ACT-R models. The parameters for the inference task were set in ACT-R so that all memory retrievals as well as retrieval times of the models matched the responses of each participant in the recognition and knowledge tasks (for details on the computations and parameter settings for memory retrieval, see Appendix B; for the full set of equations underlying ACT-R's memory, see Anderson, 2007). Note that the recognition task was administered after the inference task; this means that participants had to determine whether they had encountered the city name before or only within the experiment (i.e., a test of source memory). Therefore, the estimated recognition times in the recognition task may not provide a pure estimate of the actual recognition times in the inference task. However, in our view they still constitute the best estimates for (relative) memory accessibility and recognition times of the cities for this study.

To incorporate processes for the perception of these retrieval times, we used ACT-R's temporal module for time estimation (Taatgen, van Rijn, & Anderson, 2007). These processes provide a means to access the activation of retrieved memories and translate them into a psychological basis for decisions (see also Marewski & Schooler, 2011). Specifically, in ACT-R, time is measured internally with a noisy clock in which the length of its discrete time ticks increases over time. The models start the clock when a retrieval request is made, stop the clock when a retrieval request finishes, and store the resulting time ticks in working memory, where they are available as a basis for decisions (for the equations and parameter setting of ACT-R's temporal module, see Appendix B).

Lex-R-K evaluates the resulting retrieval time ticks against the absolute or the difference threshold to determine whether the decision can be based on recognition (time) or whether additional knowledge should be searched. How often Lex-R-K searches for knowledge depends on how these thresholds are set. Given that no information for setting these thresholds (that were defined in terms of the number of time ticks) was available prior to the modeling competition, we present Lex-R-K's predictions for three representative threshold levels. Using these levels, the three versions of Lex-R-K search for knowledge in a small, intermediate, or large number of cases, that is, about approximately 25%, 50%, or 75% of the cities that each participant recognized. The absolute thresholds were set such that these proportions were achieved for the trials of each participant in which a single city was recognized. Similarly, the difference thresholds were set such that these proportions were achieved for the trials of each participant in which both cities were recognized. Thus, the threshold values correspond to liberal, intermediate, or conservative levels in terms of how often recognition information is judged as unreliable and additional knowledge is searched. The threshold values were derived using the recognition time distributions of cities for each participant in the recognition task, used as input for the ACT-R models of Lex-R-K, but not fitted to the data of the inference task (for details see Appendix B).

3.3. Prediction of fMRI data based on the ACT-R models

Predictions for the neural data were derived following the default procedure in ACT-R (e.g., Anderson, 2007; Borst & Anderson, 2015). From the models we extracted the times when on each trial the ACT-R modules for memory retrieval and working memory updating were active. These demand functions were convolved with the hemodynamic response function as used in SPM 8 (a mix of two gamma functions; Friston, Ashburner, Kiebel, Nichols, & Penny, 2007), which resulted in an estimate for the BOLD response in the ROIs. For the retrieval ROI, only successful retrievals, not failures, contributed to the predictions, as the ROI indicates the effort of retrieving information. This corresponds to processes of a recovery phase for retrieved memories within models of associative memory (e.g., Raaijmakers & Shiffrin, 1981). For the working memory ROI, working memory updates after these successful retrievals contributed to the predictions. We extracted the predicted BOLD values at the times corresponding to the five scans of each trial and calculated the final predictions in the same way as for the observed data.

4. Results

On the basis of participants' responses in the recognition and knowledge tasks, we distinguish between six experimental conditions in the inference task (see Table 1). The conditions differ in terms of whether neither (UU), one (RU), or both (RR) cities in a pair were recognized and whether additional knowledge was present for neither (00), one (10), or both (11) recognized cities. A total of 134 of the 4185 trials were excluded from the analysis: 5 trials because of technical failure in the display of cities in the inference task, 24 trials because of technical failure in the display of cities in the recognition and knowledge tasks, 96 trials because of missed decisions, and 9 trials because of missed responses in the metacognitive judgment task.

The data were analyzed by fitting linear mixed-effects models with two regressors coding linear trends for recognition and knowledge and random effects for participants (using lme from R package nlme; Pinheiro, Bates, DebRoy, Sarkar, & The R Development Core Team, 2012; R Development Core Team, 2012; as described in Field, Miles, & Field, 2012); plots were created using ggplot2 (Wickham, 2009). To compare conditions (e.g., RU00 vs. RU10), we constructed five orthogonal contrasts from a mixed-effects model with one condition regressor.

We first report the observed response times and BOLD responses from the empirical study. Then we compare the models in terms of how well they captured the observed response time patterns, followed by a model comparison on the BOLD responses.[10] A model comparison on the choice data, which largely agreed with our conclusions from the response time and BOLD response analyses, is reported in Appendix C.

4.1. Response times

We first consider response times of participants. The top left of Fig. 3 shows the means (across participants) of the median (across trials) response times in each condition. There was an overall effect of both recognition, $\chi^2(2) = 24.439$, $p < 0.001$, and knowledge, $\chi^2(2) = 38.385$, $p < 0.001$. Orthogonal contrasts (see Table 2) indicated that response times were longer when neither of the cities was recognized than when one or both cities were recognized, $b = -34.74$, $t(128) = -4.572$, $p < 0.001$, $r = 0.37$ (i.e., UU00 vs. RU00, RU10, RR00, RR10, RR11). Moreover, response times were longer when both cities were recognized than when only one city was recognized, $b = 29.23$, $t(128) = 3.829$, $p < 0.001$, $r = 0.32$ (i.e., RR00, RR10, RR11 vs. RU00, RU10). Interestingly, response times were longer when there was no knowledge available compared to when there was knowledge available; this held both for trials with one recognized city, $b = -145.97$, $t(128) = -4.968$, $p < 0.001$, $r = 0.40$ (i.e., RU00 vs. RU10), and for trials with two recognized cities, $b = -67.08$, $t(128) = -3.835$, $p < 0.001$, $r = 0.32$ (i.e., RR00 vs. RR10, RR11). Response times did not differ between trials for

[10] Note that studies on the recognition heuristic sometimes report results separately for when participants had chosen the recognized versus unrecognized alternative (e.g., Pachur & Hertwig, 2006). However, in trials in which one city was recognized but not the other, participants in the present study chose the unrecognized city in only 8.79% of the trials (the percentages were somewhat lower in RU10 cases, 4.18%, than in RU00 cases, 14.77%). Given the low frequency of these cases, we do not report the model comparisons for response times and BOLD responses separately for which city was chosen.

Table 1
The six experimental conditions and the number of trials in the study.

Knowledge	Recognition		
	Neither city recognized (UU)	One city recognized (RU)	Both cities recognized (RR)
No knowledge about either city (00)	UU00 (814)	RU00 (738)	RR00 (263)
Knowledge about one city (10)		RU10 (958)	RR10 (560)
Knowledge about both cities (11)			RR11 (718)

Note. If both cities were unrecognized (UU), there was no knowledge available for either city (UU00). If one of two cities was recognized (RU), there was either no knowledge (RU00), or knowledge (RU10) about the recognized city. If both cities were recognized (RR), there was no knowledge about either city (RR00), or knowledge about one of the cities (RR10), or knowledge about both cities (RR11). The numbers in parentheses in each cell indicate the number of trials in that category that were included in the analysis.

which knowledge was available for one versus both of two recognized cities, $b = 36.58$, $t(128) = 1.245$, $p = 0.215$, $r = 0.11$ (i.e., RR10 vs. RR11).

It should be noted that these results by themselves are not indicative of which strategy people used because response times are sensitive not only to the amount of retrieved information (e.g., recognition vs. recognition and retrieval of additional knowledge) but also to the accessibility of each piece of information in memory that drives the associated retrieval times (e.g., Anderson, 2007). As a consequence, retrieval times (i.e., recognition and rejection times as well as retrieval and retrieval failure times for knowledge) may be different in the city pairs in different conditions. Take, for instance, our observation that response times were longer in trials in which no additional knowledge was available than in trials in which knowledge was available. On the one hand, this pattern could result from people using a knowledge-based strategy if the times for recognizing (or rejecting) city names and for searching knowledge about recognized cities are longer when no knowledge is available (and the search thus fails to retrieve anything) than when knowledge is available. On the other hand, this pattern could also result from people using a recognition-based strategy if the times for recognizing (or rejecting) city names are longer when there is no knowledge about the cities (as cities without additional knowledge may have a lower accessibility in memory) than when there is knowledge.

4.2. BOLD response in the retrieval and working memory ROIs

The BOLD response of participants in the retrieval ROI (i.e., the region around the inferior frontal gyrus in the PFC) is shown in the top left of Fig. 4. There was an effect of recognition, $\chi^2(2) = 19.460$, $p < 0.001$, and a marginal effect of knowledge, $\chi^2(2) = 5.674$, $p = 0.059$. Orthogonal contrasts (see Table 2) indicated a higher BOLD activation in trials when one or both cities were recognized than when neither of the cities was recognized, $b = 0.05$, $t(128) = 3.954$, $p < 0.001$, $r = 0.33$ (i.e., RU00, RU10, RR00, RR10, RR11 vs. UU00). Furthermore, activation was higher when both cities were recognized than when only one city was recognized, $b = 0.03$, $t(128) = 2.569$, $p = 0.011$, $r = 0.22$ (i.e., RR00, RR10, RR11 vs. RU00, RU10). Activation did not differ, however, with and without knowledge, either in trials with one recognized city, $b = 0.04$,

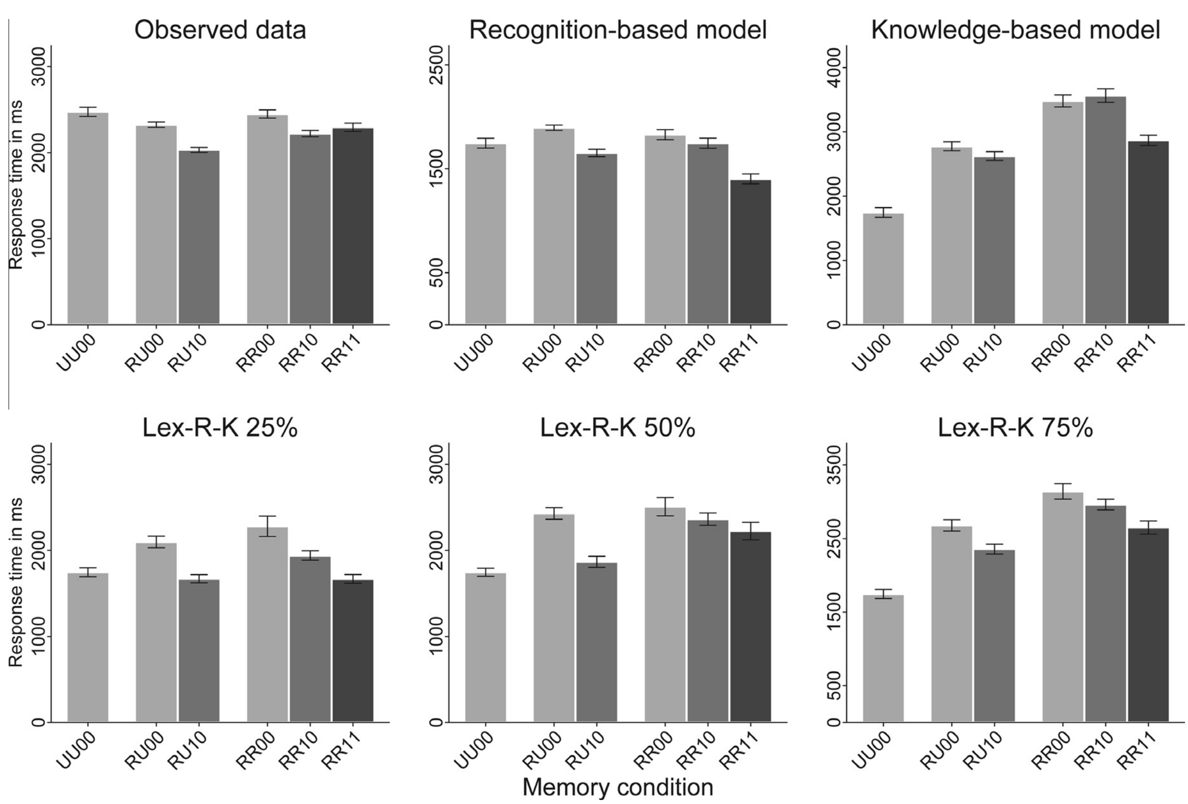

Fig. 3. Response times of participants and model predictions. Shown are the results for response times (in milliseconds) of participants (top row, left column) and the predictions of the recognition-based model (top row, middle column), the knowledge-based model (top row, right column), and the different versions of Lex-R-K that search for knowledge about 25% (bottom row, left column), 50% (bottom row, middle column), and 75% (bottom row, right column) of the cities that each participant recognized. Bars are grouped by recognition of neither (UU), one (RU), or both (RR) cities. Darker shades of gray indicate more knowledge, that is, knowledge about neither (00), one (10), or both (11) recognized cities. Error bars indicate standard errors corrected for within-subject designs (e.g., Morey, 2008). By design, the data of participants and the model predictions are shown with different limits to the axes of the dependent variables (see Section 4.3 for additional details).

Table 2
Orthogonal contrasts for differences in response times and BOLD responses.

Contrast	*b*	*t*	*df*	*p*	*r*
For response times					
UU00 vs. RU00, RU10, RR00, RR10, RR11	−34.74	−4.572	128	<0.001	0.37
RU00, RU10 vs. RR00, RR10, RR11	29.23	3.829	128	<0.001	0.32
RU00 vs. RU10	−145.97	−4.968	128	<0.001	0.40
RR00 vs. RR10, RR11	−67.08	−3.835	128	<0.001	0.32
RR10 vs. RR11	36.58	1.245	128	0.215	0.11
For BOLD response in retrieval ROI					
UU00 vs. RU00, RU10, RR00, RR10, RR11	0.05	3.954	128	<0.001	0.33
RU00, RU10 vs. RR00, RR10, RR11	0.03	2.569	128	0.011	0.22
RU00 vs. RU10	0.04	0.769	128	0.443	0.07
RR00 vs. RR10, RR11	0.04	1.419	128	0.158	0.12
RR10 vs. RR11	0.09	1.741	128	0.084	0.15
For BOLD response in working memory ROI					
UU00 vs. RU00, RU10, RR00, RR10, RR11	0.04	3.067	128	0.003	0.26
RU00, RU10 vs. RR00, RR10, RR11	0.03	1.769	128	0.079	0.15
RU00 vs. RU10	0.06	1.054	128	0.294	0.09
RR00 vs. RR10, RR11	0.06	1.776	128	0.078	0.16
RR10 vs. RR11	0.06	1.138	128	0.257	0.10

Note. This table breaks down the main effect of memory condition (i.e., UU00, RU00, RU10, RR00, RR10, and RR11) on response times and the blood-oxygen-level-dependent (BOLD) responses in the retrieval region of interest (ROI) and working memory ROI. UU = neither city recognized; RU = one city recognized; RR = both cities recognized; the notations 00, 10, and 11 indicate that knowledge was available for neither, one, or both cities, respectively.

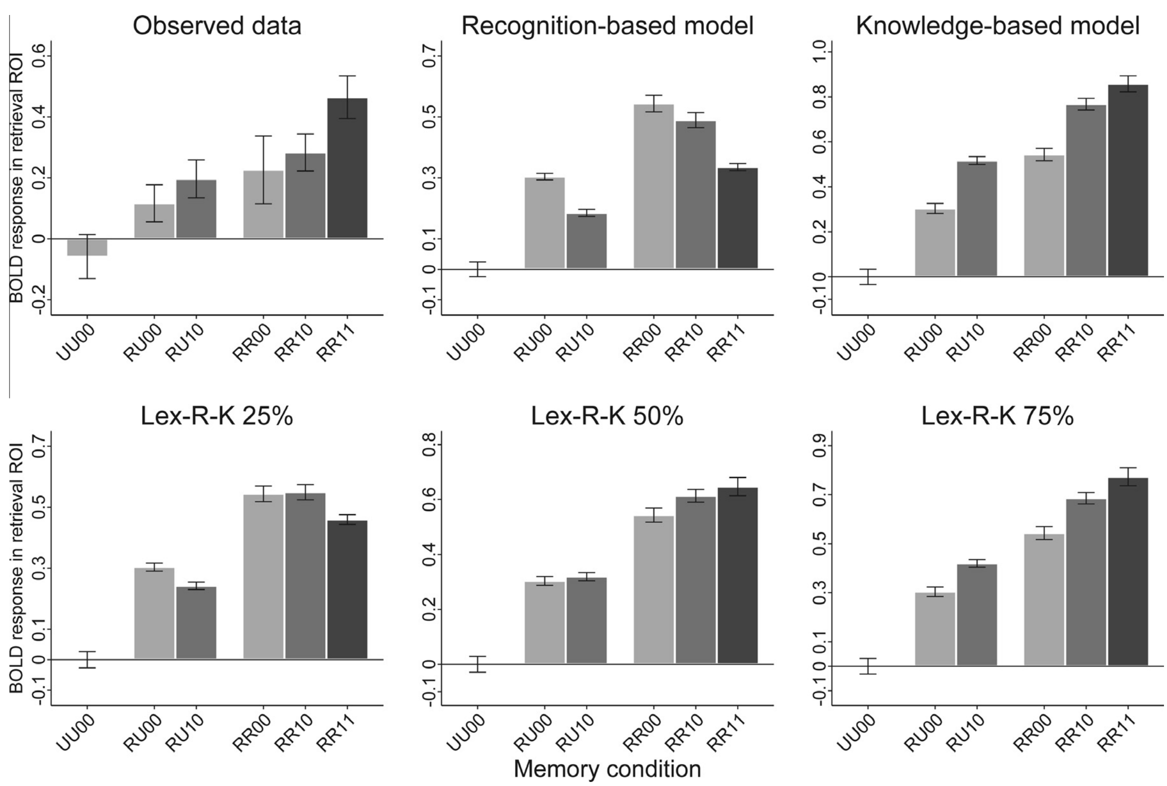

Fig. 4. BOLD response in the retrieval region of interest (ROI) of participants and model predictions. Shown are the results for the blood-oxygen-level-dependent (BOLD) response in the retrieval ROI of participants (top row, left column) and the predictions of the recognition-based model (top row, middle column), the knowledge-based model (top row, right column), and the different versions of Lex-R-K that search for knowledge about 25% (bottom row, left column), 50% (bottom row, middle column), and 75% (bottom row, right column) of the cities that each participant recognized. Bars are grouped by recognition of neither (UU), one (RU), or both (RR) cities. Darker shades of gray indicate more knowledge, that is, knowledge about neither (00), one (10), or both (11) recognized cities. Error bars indicate standard errors corrected for within-subject designs.

$t(128) = 0.769$, $p = 0.443$, $r = 0.07$ (i.e., RU10 vs. RU00), or in trials with two recognized cities, $b = 0.04$, $t(128) = 1.419$, $p = 0.158$, $r = 0.12$ (i.e., RR10, RR11 vs. RR00). There was a tendency for activation to be slightly higher in trials with knowledge about both recognized cities compared to knowledge about only one of them, $b = 0.09$, $t(128) = 1.741$, $p = 0.084$, $r = 0.15$ (i.e., RR11 vs. RR10).

The BOLD response of participants in the working memory ROI (i.e., the region close to the intraparietal sulcus in the PPC) is shown in the top left of Fig. 5. Again, there was an effect of recognition, $\chi^2(2) = 12.100$, $p = 0.002$, and a tendency toward an effect of knowledge, $\chi^2(2) = 5.655$, $p = 0.059$. Activation was higher in trials in which one or both cities were recognized compared to when

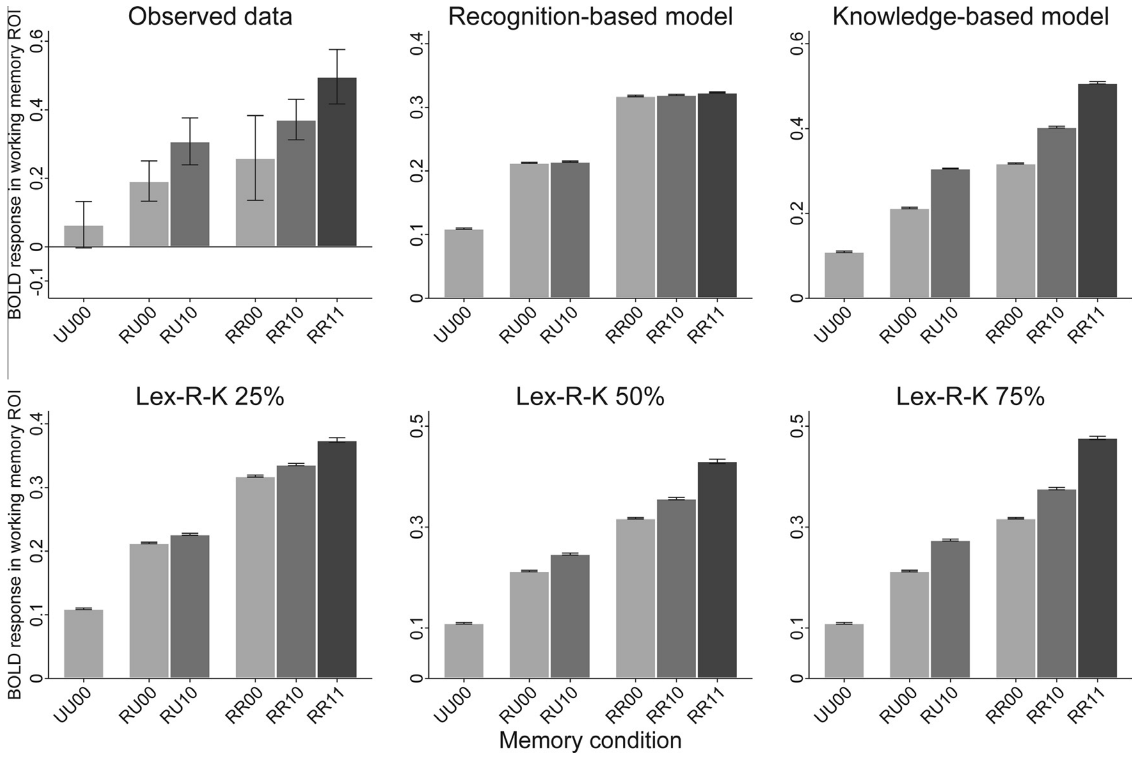

Fig. 5. BOLD response in the working memory region of interest (ROI) of participants and model predictions. Shown are the results for the blood-oxygen-level-dependent (BOLD) response in the working memory ROI of participants (top row, left column) and the predictions of the recognition-based model (top row, middle column), the knowledge-based model (top row, right column), and the versions of Lex-R-K that search for knowledge about 25% (bottom row, left column), 50% (bottom row, middle column), and 75% (bottom row, right column) of the cities that each participant recognized. Bars are grouped by recognition of neither (UU), one (RU), or both (RR) cities. Darker shades of gray indicate more knowledge, that is, knowledge about neither (00), one (10), or both (11) recognized cities. Error bars indicate standard errors corrected for within-subject designs.

neither city was recognized, $b = 0.04$, $t(128) = 3.067$, $p = 0.003$, $r = 0.26$ (i.e., RU00, RU10, RR00, RR10, RR11 vs. UU00; see Table 2). There was a tendency for activation to be higher when both cities were recognized than when only one city was recognized, $b = 0.03$, $t(128) = 1.769$, $p = 0.079$, $r = 0.15$ (i.e., RR00, RR10, RR11 vs. RU00, RU10). Activation did not differ with and without knowledge in trials in which only one city was recognized, $b = 0.06$, $t(128) = 1.054$, $p = 0.294$, $r = 0.09$ (i.e., RU10 vs. RU00). For trials with two recognized cities there was a tendency for a higher activation when there was knowledge about one or both cities than when there was no knowledge, $b = 0.06$, $t(128) = 1.776$, $p = 0.078$, $r = 0.16$ (i.e., RR10, RR11 vs. RR00), but activation did not differ between trials with knowledge for one versus both of two recognized cities, $b = 0.06$, $t(128) = 1.138$, $p = 0.257$, $r = 0.10$ (i.e., RR10 vs. RR11).

Note again that the neural measures, like the response times, are sensitive not only to the amount (i.e., recognition and knowledge, or only recognition) but also to the accessibility of memories. They are not, on their own, indicative of whether a recognition-based or knowledge-based strategy was used. In the next section, we therefore compare the response times and neural patterns to the model predictions of the different strategies, which take into account both accessibility and the amount of retrieved information.

4.3. Test of the model predictions of response times

First, we examine the response time predictions of the models that are shown in Fig. 3. Our initial analysis of these predictions resulted in very poor fits of all models to the data. An inspection of the underlying response time distributions as shown in Appendix D revealed that this was mainly due to all models underestimating response times in the UU00 condition, that is, when neither of the decision alternatives was recognized. Therefore, we report the numerical fit of the models with and without the UU00 condition, which is presented in Table 3. One potential reason for the poor model performance in the UU00 condition is that after failing to recognize the alternatives, all models simply guess; participants, in contrast, may engage in strategies not included in our models (for a discussion of such potential strategies, see Appendix D).

We start with the predictions of the recognition-based and the knowledge-based model, which are shown in the middle and right columns of the top row in Fig. 3.[11] Across conditions, neither the recognition-based nor the knowledge-based model adequately accounted for the observed pattern of response times. The recognition-based model underestimated the observed response times especially when both cities were recognized and knowledge was available for both of them, compared to when there was knowledge available about only one or neither of the cities (i.e., RR11 vs. RR10, RR00). This may indicate that, in contrast to the strong assumption of the recognition-based model that knowledge beyond recognition is never searched (which produces relatively fast

[11] In the figures, the data of participants and the model predictions are shown with different limits to the axes of the dependent variables. This was done to emphasize that the absolute values of the model predictions are not necessarily informative about how well they account for the data because the values can overall be scaled for each model and dependent measure (within a limited range). In contrast, the differences between the experimental conditions for each model are more informative and are discussed when we evaluate the models against the data. As a consequence, we also use r as an index of model performance (in addition to $RMSD$) as it allows us to evaluate how well the models approximate the data independent of scale.

Table 3
Indices for the correspondence between the model predictions and the observed data.

Model	Response time		BOLD response in retrieval ROI		BOLD response in working memory ROI	
	RMSD	*r*	*RMSD*	*r*	*RMSD*	*r*
Without UU00						
Recognition-based model	588.61	0.30	0.20	0.15	0.10	0.62
Lex-R-K model 25%	369.03	0.77	0.20	0.48	0.07	0.80
Lex-R-K model 50%	115.21	0.91	0.24	0.83	0.05	0.90
Lex-R-K model 75%	525.65	0.79	0.30	0.93	0.03	0.95
Knowledge-based model	867.10	0.51	0.35	0.93	0.03	0.98
With UU00						
Recognition-based model	614.05	0.31	0.18	0.63	0.09	0.83
Lex-R-K model 25%	449.24	0.46	0.19	0.79	0.07	0.91
Lex-R-K model 50%	315.27	0.19	0.22	0.93	0.05	0.95
Lex-R-K model 75%	564.44	−0.08	0.27	0.97	0.04	0.98
Knowledge-based model	845.51	−0.18	0.32	0.97	0.03	0.99

Note. The indices (*RMSD* and *r*) refer to the correspondence between the aggregate (across participants) model predictions and aggregate (across participants) data. BOLD = blood-oxygen-level-dependent; ROI = region of interest; UU00 = neither city was recognized and there was no knowledge available for either city; Lex-R-K model 25%, 50%, or 75% = versions of Lex-R-K that search for knowledge about 25%, 50%, or 75% of the cities that each participant recognized.

response times), participants do consider additional knowledge—at least sometimes.

The knowledge-based model, by contrast, predicted overall considerably longer response times when both cities were recognized than when only one city was recognized (i.e., RR00, RR10, RR11 vs. RU00, RU10); although the empirical data indeed showed such a difference, this difference was less pronounced than was predicted by the model (see also the response time distributions for RR vs. RU conditions in Appendix D). The knowledge-based model thus seems to err in the assumption that a search for additional knowledge is initiated for all recognized cities. When both cities were recognized, the knowledge-based model predicted longer response times when knowledge was available about neither or one city than when knowledge was available about both cities (i.e., RR00, RR10 vs. RR11); also this difference was overestimated by the model compared to the observed data. Again, one reason could be that the model overestimates how often people search for knowledge in cases where knowledge about neither or one city was available.

On the other hand, both the recognition-based and the knowledge-based model correctly predicted that in trials with one recognized city (and for the recognition-based model also in trials with two recognized cities) the response times were longer when no knowledge was present compared to when there was knowledge (i.e., RU00 vs. RU10; RR00 vs. RR10, RR11). In the recognition-based model this results from slower recognition times for cities without knowledge compared to cities with knowledge. In the knowledge-based model, retrieval failures for knowledge on these slowly recognized cities add to the effect.

The predictions of the different versions of Lex-R-K for response times are shown in the bottom row of Fig. 3. In general, for all three versions the predictions matched the observed pattern of response times better than those of the recognition-based and the knowledge-based model, with the 50% version of Lex-R-K providing the best account (i.e., in both *RMSD* and *r* without considering the UU00 condition; see also the response time distributions in Appendix D). This indicates that in some but not all cases, knowledge beyond recognition was considered. To explore if these results also hold for the data of individual participants, we examined the model performance for individual response times (see Appendix E).

4.4. Test of the model predictions of the BOLD response

Next, we examine the model predictions of the BOLD response. Given that the models do not predict the exact size of the increase of the BOLD response, we focus on *r* for assessing the model performance as it is independent of scale. The results for the retrieval ROI (i.e., the region around the inferior frontal gyrus in the PFC) are shown in Fig. 4 and the results for the working memory ROI (i.e., the region close to the intraparietal sulcus in the PPC) in Fig. 5. For the UU00 condition, all model predictions were in line with the observed BOLD response, which showed no increase in either of the two ROIs. This indicates that the processes that prolong response times of participants in the UU00 condition (Fig. 3) do not map on activity in the retrieval or working memory ROI. For consistency, we nevertheless provide the numerical fit of the models with and without the UU00 condition in Table 3.

For the BOLD response in the retrieval ROI, the predictions of the recognition-based and the knowledge-based model are shown in the middle and right columns of the top row in Fig. 4. Overall, the knowledge-based model matched the observed pattern better than the recognition-based model. For trials with one or two recognized cities the recognition-based model predicted a lower activation in trials with more knowledge, whereas the observed data showed a tendency for higher activation with more knowledge. The reason for this prediction of the recognition-based model is that cities with associated knowledge tend to be retrieved faster. This resulted in less demand on the retrieval module. The better performance of the knowledge-based model compared to the recognition-based model suggests that participants do consider retrieving knowledge beyond recognition in at least some of the decisions.

The predictions of the different versions of Lex-R-K for the BOLD response in the retrieval ROI are shown in the bottom row of Fig. 4. The predictions of all three versions matched the pattern of the fMRI data better than those of the recognition-based model. In particular the 50% and the 75% versions of Lex-R-K matched the observed data pattern rather well, predicting an additional increase of the BOLD response when knowledge about more cities could be retrieved. These versions of Lex-R-K provided a similarly good account of the data as the knowledge-based model.

For the working memory ROI, the predictions of the recognition-based and the knowledge-based model are shown in the top row, middle and right columns in Fig. 5. Here, again the knowledge-based model accounted for the observed data better than the recognition-based model. The main reason is that the knowledge-based model predicted the observed tendency for activation to increase with knowledge about more cities. In contrast, the recognition-based model predicted no increase in activation with increasing knowledge. This provides further support for the assumption that knowledge is retrieved and stored in working memory in at least some of the decisions.

The predictions of the different versions of Lex-R-K for the working memory ROI can be found in the bottom row of Fig. 5. The predictions of all versions of Lex-R-K matched the pattern of the BOLD response better than those of the recognition-based model, predicting a higher activation when knowledge about more cities could be retrieved. Again, versions of Lex-R-K that searched for knowledge about 50% or 75% of the recognized cities accounted well for the observed data pattern and provided a similarly good account of the data as the knowledge-based model.

As mentioned above, in Appendix C we report a model comparison using participants' choices, making the simplifying assumption that for the trials in which a choice was based on knowledge, the city for which knowledge was retrieved (or when there was knowledge about both cities, the city for which knowledge was retrieved faster) was inferred to be larger. These analyses similarly favored the 50% and 75% versions of Lex-R-K and the knowledge-based model over the recognition-based model and the 25% version of Lex-R-K.

5. General discussion

We investigated how people make decisions based on their real-world memories, using a task in which people judged which of two cities has more inhabitants. We tested a recognition-based strategy, a knowledge-based strategy, and the strategy Lex-R-K, which sequentially considers recognition and—when recognition is deemed an unreliable basis for a decision—moves on to consider additional knowledge. We developed computational models for these strategies in ACT-R that took into account the accessibility of individual memories (in terms of retrieval times) as well as cognitive processes for perceiving these retrieval times. Lex-R-K was implemented and tested in three versions, that searched for additional knowledge about a small, intermediate, and large number of recognized alternatives. We then evaluated the models in terms of how well they matched the observed patterns of response times and BOLD activations; here we focus on the conclusions from the conditions when one or both decision alternatives were recognized.

The main finding of our study is that the different versions of Lex-R-K, which search for additional knowledge only when recognition information is deemed unreliable, were better able to account for the joint patterns of behavioral and neural data than models that either always consider only recognition information or always extend information search to additional knowledge. As regards the response times, the predictions of the different versions of Lex-R-K were more consistent with the observed patterns than the predictions of the recognition-based and the knowledge-based model, which seemed to underestimate and overestimate, respectively, how often participants searched for knowledge (Fig. 3); the version of Lex-R-K that searched for knowledge about an intermediate proportion of the recognized alternatives showed the closest correspondence to the data.

As regards the fMRI data, all versions of Lex-R-K and the knowledge-based model provided a better account of the activation in the retrieval ROI in the PFC and working memory ROI in the PPC than the recognition-based model. This was particularly the case where these models, but not the recognition-based model, predicted an increase in the BOLD response when additional knowledge about recognized alternatives could be retrieved (Figs. 4 and 5). In contrast to the response time data, for the BOLD data the knowledge-based model gave a similarly good account as the best-performing versions of Lex-R-K. One possible explanation for the finding that the knowledge-based model performed better on the BOLD data than on the response time data is that participants sometimes might have engaged in postdecisional information search (for knowledge) in memory (cf. Pleskac & Busemeyer, 2010). This would be reflected in the BOLD data (which are recorded even after a decision has been made) but not in the response times.

Overall, we find that a strategy according to which knowledge is searched conditionally on the evidence obtained from recognition memory approximated the observed joint data patterns of response times and BOLD responses best. Thus, rather than strategies that blindly rely on either recognition or knowledge, a strategy that evaluates the available recognition information on whether it gives a reliable basis for decisions described participants' behavior well; only in the case of insufficient recognition information does it engage in further cognitive processing to search for knowledge. Such a strategy exploits the available information as much as necessary to make decisions while at the same time using cognitive resources in a parsimonious way. On a more general note, our findings highlight that additional knowledge beyond recognition is usually not ignored in memory-based decision making (as is sometimes assumed; e.g., Goldstein & Gigerenzer, 2002). However, they also show that a search for knowledge in memory is not always initiated but depends on the evidence from recognition memory that has already been obtained. Future research should explore in greater detail the specifics of this conditional retrieval of additional knowledge and how the knowledge is evaluated when making a decision.

For evaluating the superior performance of Lex-R-K in the model comparison, it should first be emphasized that none of the models tested had any free parameters that were fitted to the data of the inference task. The models were adjusted to the individual participants, but this was based on data collected in the separate recognition and knowledge tasks, not the inference task. Furthermore, it is important to note that the recognition-based and knowledge-based models were not fully nested under Lex-R-K, as the models differ qualitatively in their production rules. Whereas the models overlap in production rules for perceiving and recognizing the alternatives and giving a motor response, the models differ in the production rules that evaluate recognition information and initiate knowledge search.

In contrast to the recognition-based and knowledge-based strategies, Lex-R-K relied on thresholds to evaluate recognition information. In principle, this could have given the model more flexibility in fitting the data. To acknowledge this potential advantage of Lex-R-K, however, we tested it in three versions with three a priori defined fixed threshold levels. These levels were chosen such that the tested versions of Lex-R-K would predict qualitatively different patterns than the recognition-based and the knowledge-based model. As a consequence, the superiority of the three versions of Lex-R-K is not due to them being more flexible than the other two models. Had participants relied on the recognition-based or the knowledge-based strategy, the models of these strategies would have outperformed the different versions of Lex-R-K.

Further, note that all three versions of Lex-R-K generally captured response times better than the other two models (Table 3), so Lex-R-K's superiority is not due to the fact that three versions of it entered the model comparison. For the BOLD response, all versions of Lex-R-K provided a similarly good account of the data in the working memory ROI as the knowledge-based model; only for the retrieval ROI did the 25% version of Lex-R-K provide a less good account of the data than the 50% and 75% versions of Lex-R-K and the knowledge-based model. In what follows, we discuss the implications of our findings.

5.1. Toward a better understanding of the contingent use of recognition and knowledge in decision making

The results of our study suggest that when people make inferences about real-world objects, recognition and additional knowledge may be recruited depending on the particular states of memories. The Lex-R-K strategy evaluates the reliability of the recognition information before using it for a decision; only when recognition information is judged to be unreliable does it go on to retrieve additional knowledge. When implementing Lex-R-K in the ACT-R architecture, we assumed that the reliability of recognition information is judged by exploiting the abilities for estimating and evaluating recognition times (cf. Taatgen et al., 2007; see also Marewski & Schooler, 2011). This implementation thus specifies the cognitive processes that enable the execution of Lex-R-K.

Whereas most existing models of memory-based decision making assume that either all available knowledge or only recognition (and/or fluency) is used for decision making, Lex-R-K is a first formal modeling account of the interplay between these different types of information at the level of cognitive processing. The support for the mechanism in Lex-R-K also suggests how recognition speed might act as a "gatekeeper" for whether to search for additional information. That our implementation of the models in ACT-R also provides predictions for neural activation further helps enrich the empirical content of the model predictions. It illustrates how a multimethod approach to model testing and model classification that includes neuroimaging data can be used to extend our understanding of the cognitive processes in decision making.

The support for Lex-R-K does not rule out that in principle other strategies might also be able to account for participants' behavior when making inferences based on recognition or knowledge. Whereas the current study has focused on mechanisms implementing serial information search, future studies could also consider mechanisms such as exemplar-based strategies (Juslin, Olsson, & Olsson, 2003; Pachur & Olsson, 2012) and implement them in ACT-R to assess the circumstances under which strategies with serial and parallel search for information are employed in memory-based inferences. Furthermore, none of the models tested here provided a good account of the data in the condition in which neither of the possible alternatives was recognized; specifically, all models systematically underestimated response times. It is possible that rather than simply guessing in these cases (as assumed by the models) people repeatedly attempted and failed to retrieve information associated with the alternatives. Future studies could examine in greater detail the cognitive processes when people make inferences about completely unknown alternatives.

In addition, further research should investigate in more detail the process by which recognition information (as well as other information from memory) is evaluated and clarify the conditions under which particular aspects of the outcome of such evaluation processes inform memory-based inferences (for a proposal that also incorporates the evaluation of recognition quality for unrecognized alternatives, see Erdfelder et al., 2011; for a theory-integration approach based on threshold principles from the study of perception and preferences with inferences, see Luan et al., 2014). Furthermore, whereas our results were obtained for a specific real-world domain and a paired-comparison task, future research should examine to what extent our results generalize to other domains and situations with a greater number of decision alternatives.

5.2. Contribution to the strategy selection problem

Lex-R-K can be considered a mechanism for how to select between strategies, which are either recognition-based or knowledge-based. Different approaches have been proposed in the literature on how people can select between applicable strategies that exploit the available information in different ways. However, these approaches often face limitations when applied to real-world domains. Marewski and Schooler (2011) mapped out situations when recognition-based and knowledge-based strategies could be applied in real-world domains and when each kind of strategy would be most likely to lead to good outcomes. Their analysis, however, left open the question of how the cognitive system could select the most appropriate strategy. One possible answer would be a meta-strategy that weighs the costs and benefits associated with each strategy (e.g., Payne et al., 1993). It is often not the case that such costs and benefits can be known in real-world decision situations. Selection could also be tuned by experience, for instance, via reinforcement learning in the presence of outcome feedback (e.g., Rieskamp & Otto, 2006); but, again, such immediate feedback is often unavailable in real decision situations. Lex-R-K provides our answer to the question of how to select a strategy in the absence of pertinent cost and benefit information and immediate outcome feedback. Lex-R-K turns inward, relying on information that results from general memory-retrieval processes, shaped by experience and executed in the course of decision making. Its selection mechanism adds to the explanation of how in real-world situations—where information is often incomplete—an adaptive selection between strategies could be accomplished by the human mind.

6. Conclusion

We demonstrated how computational models can be used to formalize decision strategies and tested them based on behavioral and neuroimaging data. These computational models incorporate and specify the cognitive processes underlying decision strategies, including the dynamics of memory retrieval and the perception of retrieval times. These models made qualitative and quantitative predictions for behavioral and fMRI data that we used to evaluate the candidate strategies. The results contribute to a better understanding of how people can flexibly recruit recognition and additional knowledge from memory when making inferences about unknown properties in the real world.

Acknowledgments

This work was supported by a grant from the German Academic Exchange Service (DAAD), the Center for Adaptive Behavior and Cognition (ABC) and the Center for Adaptive Rationality (ARC) at the Max Planck Institute for Human Development, the International Max Planck Research Network on Aging (MaxNetAging), and the Werner Reichardt Centre for Integrative Neuroscience (CIN) at the Eberhard Karls University of Tübingen. The CIN is an Excellence Cluster funded by the Deutsche Forschungsgemeinschaft (DFG) within the framework of the Excellence Initiative (EXC 307). We thank John R. Anderson and his research group for hosting the first author at Carnegie Mellon University, where substantial proportions of the data analysis and computational modeling were conducted. We thank Anita Todd for editing the manuscript.

Appendix A. Flowcharts of ACT-R models for the decision strategies

Appendix A contains the flowcharts of the ACT-R models for the recognition-based strategy (Fig. A1), the knowledge-based strategy (Fig. A2), and the Lex-R-K strategy (Fig. A3).

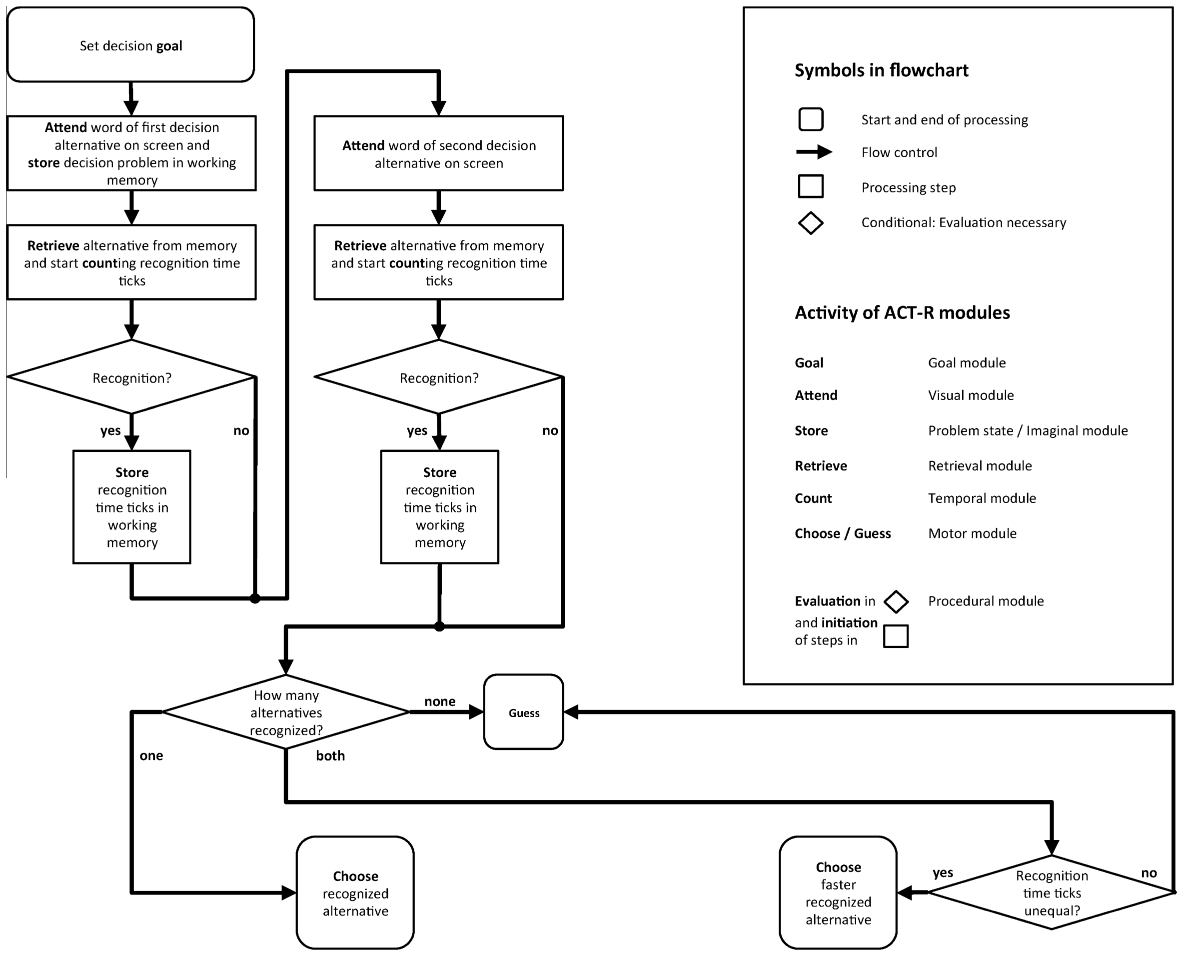

Fig. A1. Processing steps of the ACT-R model for the recognition-based strategy, implemented by the interaction of the different ACT-R modules.

Appendix B. Implementation of memory retrieval in ACT-R

This Appendix describes the dynamics of memory retrieval in ACT-R, as well as the processes for perceiving retrieval times.

B.1. Contents of ACT-R's declarative memory

To populate ACT-R's declarative memory with memory chunks of cities and knowledge facts about cities we used the data of each participant from the recognition and knowledge tasks, such that the models' memories corresponded to the memories indicated by each participant. This means that we created city and knowledge chunks that held the information to be known or unknown by each participant and read these chunks into ACT-R's declarative memory when the models were working on the trials of the inference task of this participant.

B.2. Time dynamics of memory retrieval in ACT-R

To model the time dynamics of memory retrieval for each of these memory chunks, we estimated the recognition or rejection time for each city chunk from the recognition task and the successful retrieval or retrieval failure time for each knowledge chunk from the knowledge task of the respective participant. We refer to these four kinds of times as *retrieval times*; they were obtained by subtracting nonretrieval time from the overall reaction times in the recognition and knowledge tasks.

$$retrieval\ time = reaction\ time - nonretrieval\ time \tag{B.1}$$

Nonretrieval time was estimated as 445 ms based on ACT-R standard times, consisting of visual processes (i.e., 50 ms to request and 85 ms to shift visual attention to a city word at a known location), retrieval preparation (i.e., 50 ms to request a retrieval), and the motor response (i.e., 50 ms for requesting, 150 ms for preparation, 50 ms for initiation, and 10 ms for execution of a punch movement), and kept constant across tasks, participants, and models in this study. We truncated the resulting retrieval times at ±2 *SD* away from the means of the retrieval and rejection times for cities and successful retrieval and retrieval failure times for knowledge of each participant to avoid outliers and we set resulting negative and extremely short retrieval times to short positive times (i.e., 1 ms). Fig. B1 shows the resulting retrieval times for the recognition task (top left panel) and the knowledge task (top right panel).

B.3. Memory activation in ACT-R

We transformed these retrieval times into activation levels for the chunks in memory, because in ACT-R activation is the central currency that determines the speed of memory retrieval. The retrieval time of a chunk is given by

$$retrieval\ time = Fe^{-A_i} \tag{B.2}$$

where the retrieval time of chunk i becomes shorter the higher its activation A_i is. The scaling factor F was set to 1, such that the resulting retrieval times matched the times estimated from the recognition and knowledge tasks. We obtained the activation level of each chunk by transforming Eq. (B.2) into

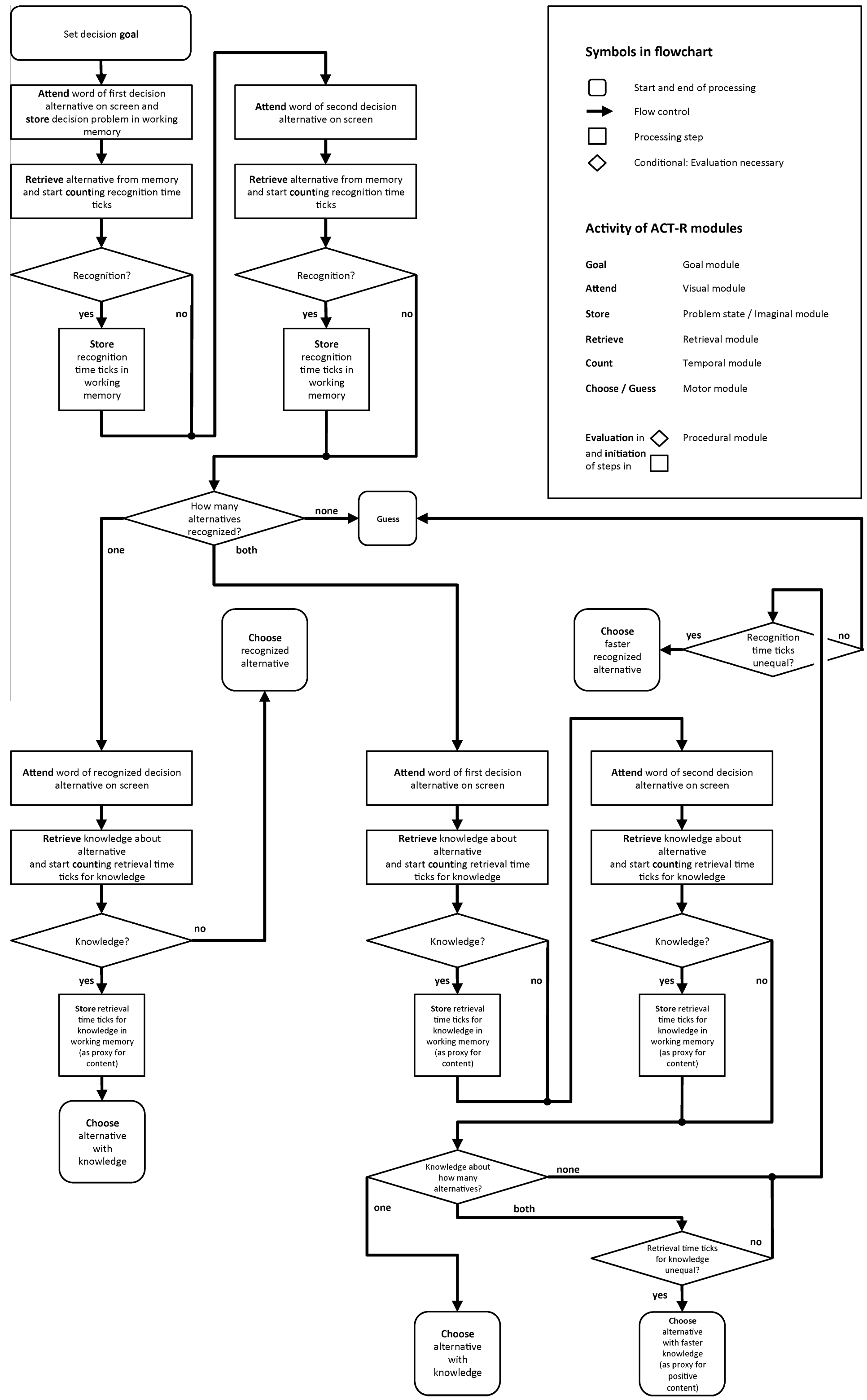

Fig. A2. Processing steps of the ACT-R model for the knowledge-based strategy, implemented by the interaction of the different ACT-R modules.

$$A_i = -ln\left(\frac{retrieval\ time}{F}\right) \tag{B.3}$$

In ACT-R, chunks can be retrieved only when their activation is above the retrieval threshold τ. Because we estimated activation from the recognition and knowledge tasks, we set τ to such a low value (i.e., -3) that all chunks from declarative memory could be retrieved. Likewise, activation noise was switched off. As a result, all chunks had exactly the activation and resulting retrieval time that we estimated from the recognition and knowledge tasks.

B.4. Perception of retrieval times in ACT-R

To enable the models to perceive and process the retrieval times of memories we relied on ACT-R's temporal module that can

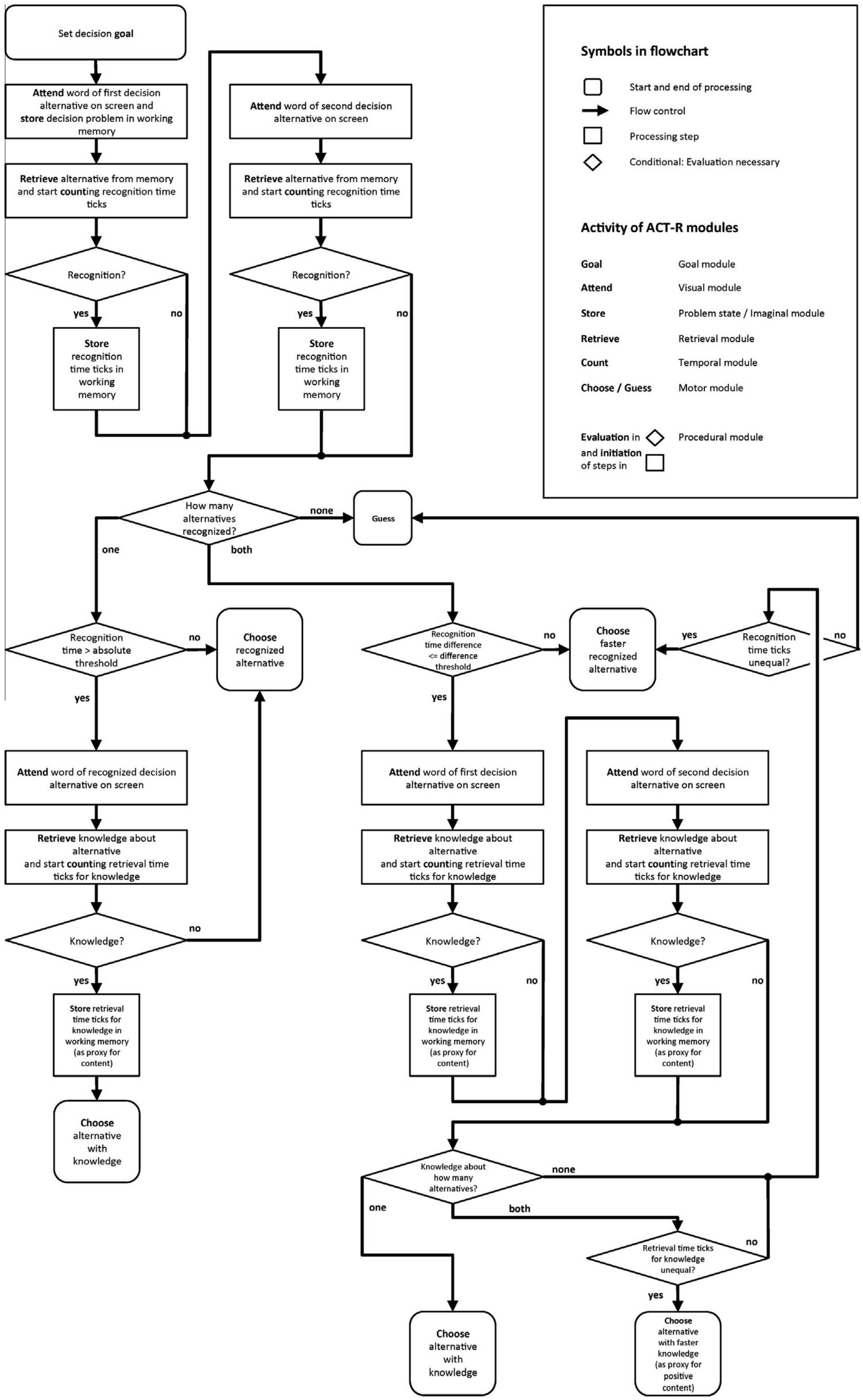

Fig. A3. Processing steps of the ACT-R model versions for the Lex-R-K strategy, which in a lexicographic way (Lex) searches for recognition (R) and knowledge (K), implemented by the interaction of the different ACT-R modules.

determine time intervals. It works like an internal timer that can count the number of ticks that have passed between when it is started and when it is stopped. The length of the first tick t_0 is given by

$$t_0 = start + \varepsilon_1 \tag{B.4}$$

where *start* defaults to 11 ms, and noise ε_1 is added (drawn from a logistic distribution with a mean of 0 and a variance σ^2 for which *s* can be set such that $\sigma^2 = \pi^2/3 * s^2$; by default generated with an *s* value of $b * 5 * start$, with *b* defaulting to 0.015). The length of the following *n*th tick is given by

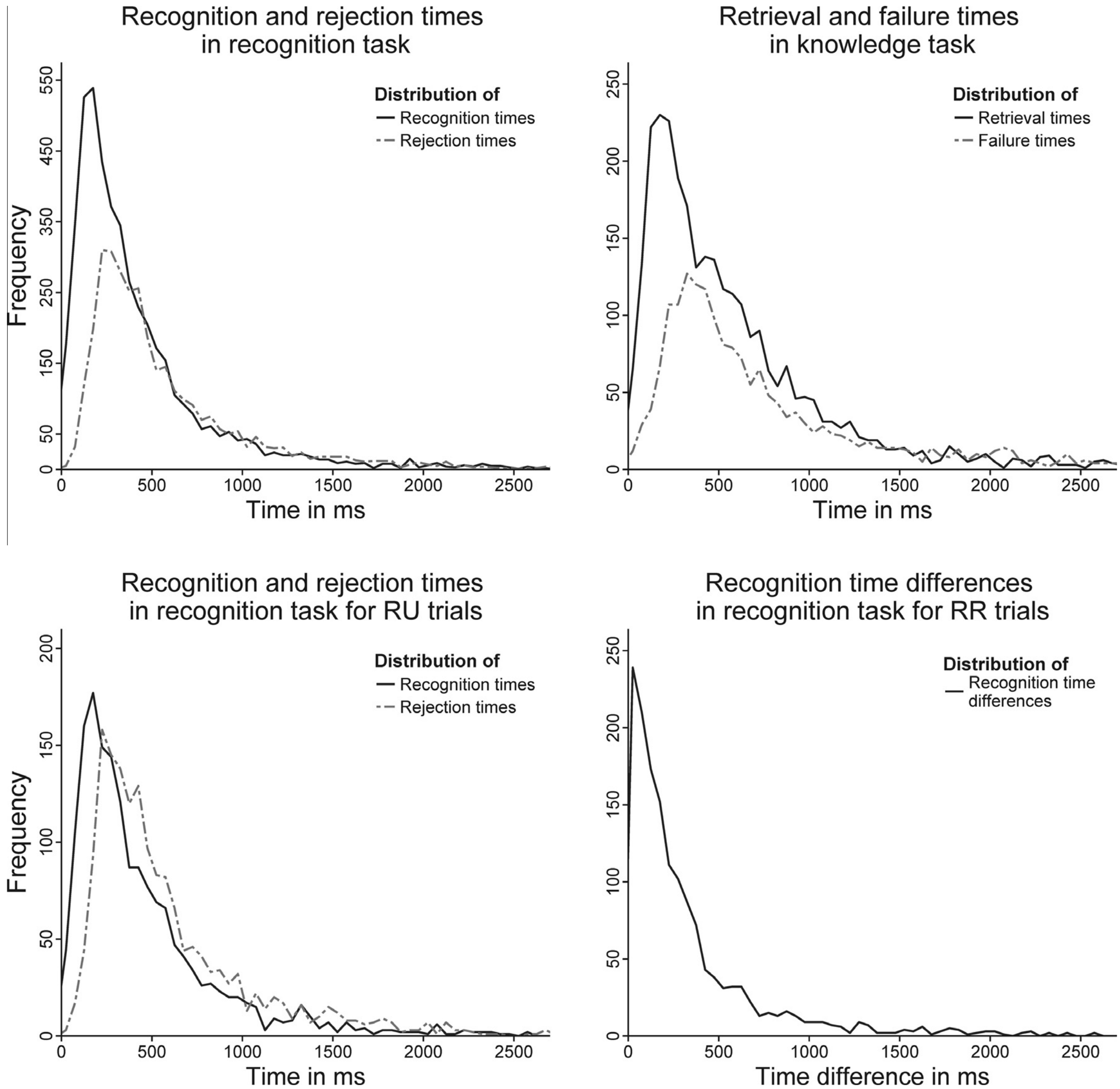

Fig. B1. Observed retrieval time distributions from the recognition and knowledge tasks. The top left panel shows the distributions of the recognition and rejection times in the recognition task, the top right panel shows the distributions of successful retrieval and retrieval failure time in the knowledge task. The bottom left panel shows the distributions of recognition and rejection time in the recognition task for trials in which only one city was recognized (i.e., RU) and the different versions of Lex-R-K apply an absolute threshold for evaluating recognition times. The bottom right panel shows the distribution of recognition time differences in the recognition task for trials where both cities were recognized (i.e., RR) and the different versions of Lex-R-K apply a difference threshold for evaluating recognition time differences. For display purposes retrieval time distributions are shown aggregated across participants with frequencies computed for bin widths of 50 ms and cut off at 2,500 ms.

$$t_n = a * t_{n-1} + \varepsilon_2 \quad \text{(B.5)}$$

where a defaults to 1.1, and noise ε_2 is added (by default generated with an s value of $b * a * t_{n-1}$). These ticks are not only noisy but also prolonged with time such that time interval estimation gets less accurate as time increases. For the present study, we relied on the default values of the temporal module except for the added noise, where we set b to a value small enough (i.e., 10^{-10}) not to affect performance differences between the tested models (as the temporal module does not accept a b of 0). When a retrieval request is made by a model, the timer of the temporal module starts counting and stops when the retrieval process finishes (i.e., the retrieved chunk is placed in the retrieval buffer). The result of this time estimation process (in the form of the measured number of ticks) is stored in ACT-R's working memory, which by default takes 200 ms, and is after that available as a basis for decisions.

B.5. Thresholds for perceived retrieval times in the different versions of Lex-R-K

The perceived retrieval time ticks that result when an attempt to recognize a city is made and that are stored in working memory are evaluated by Lex-R-K to determine whether a decision can be based on recognition (time) or if additional knowledge should be searched. If only one of two cities was recognized, Lex-R-K evaluates whether this city was recognized too slowly (compared to an absolute threshold); if both cities were recognized, Lex-R-K checks whether the recognition time difference between the cities was too small (in comparison to a difference threshold). In both of these cases, Lex-R-K moves on to search for knowledge. How often Lex-R-K searches for knowledge depends on how these thresholds are set by a person or a model when applying Lex-R-K.

For this study, we present the results of Lex-R-K for three different threshold levels that make the three versions of Lex-R-K search for knowledge for about approximately 25%, 50%, or 75% of the cities that each participant recognized. Thus, the threshold values were set to a liberal, intermediate, or conservative level in terms of how often recognition information is judged as unreliable and additional knowledge is searched. The exact threshold values were derived individually from the recognition time distributions of each participant in the recognition task. When only one city was recognized, this means that a participant (or model version) would set the absolute threshold such that she or he would search for knowledge about the 25%, 50%, or 75% of the cities recognized most slowly (see the recognition time distribution here aggregated across participants in the bottom left panel of Fig. B1). When both cities were recognized, this means that a participant (or model version) would set the difference threshold such that she or he would search for knowledge about 25%, 50%, or 75% of pairs of recognized cities for which the recognition time differences were shortest (see the recognition time difference distribution here aggregated across participants in the bottom right panel of Fig. B1). The threshold values of each participant were set in the versions of Lex-R-K before running them on the trials of the inference task of the respective participant.

Appendix C. Choices in the inference task

In this section, we present the choice data from the inference task. Fig. C1 shows the correspondence of the model predictions to the choices of participants. Overall, the correspondence patterns of the models across the memory conditions appeared very similar, which is in line with previous findings that recognition-based and knowledge-based strategies often predict the same choices (e.g., Hilbig & Pohl, 2009; Pachur et al., 2011). Specifically, in the current study all models showed a greater correspondence to the choices of participants when knowledge was available about one of the two cities compared to when there was no knowledge (i.e., RU10 vs. RU00 and RR10 vs. RR00; by relying on recognition time for the cities or on knowledge about the cities). However, all models showed a lower correspondence in cases of two recognized cities with knowledge available for both of them compared to cases with knowledge about only one of the two cities (i.e., RR11 vs. RR10). A possible reason is that no information on the content of the knowledge used in a particular decision was available to the models that could make the models choose the same alternative as participants did.

Concerning differences between the models, only when both cities were recognized and knowledge was available about one of them (i.e., RR10) did the recognition-based model and the 25% version of Lex-R-K appear to correspond to participants' choices less often than the other models. Mixed-effects models corroborated this difference. There was an effect of recognition, $\chi^2(2) = 132.401$, $p < 0.001$, knowledge, $\chi^2(2) = 69.602$, $p < 0.001$, and model, $\chi^2(4) = 16.628$, $p = 0.002$. Moreover, there was an interaction between recognition and model, $\chi^2(8) = 18.967$, $p = 0.015$, and between knowledge and model, $\chi^2(8) = 61.604$, $p < 0.001$. This corresponds to an effect of memory condition, $\chi^2(5) = 216.688$, $p < 0.001$, an effect of model, $\chi^2(4) = 16.628$, $p = 0.002$, and an interaction between memory condition and model, $\chi^2(20) = 121.821$, $p < 0.001$. Orthogonal contrasts were used to break down this interaction (Table C1). The recognition-based model showed a lower correspondence to choices of participants than the other models when both cities were recognized compared to when only one city was recognized (i.e., RR00, RR10, RR11 vs. RU00, RU10), when there was knowledge available about one or both cities compared to when there was no knowledge available (i.e., RR10, RR11 vs. RR00), and when there was knowledge

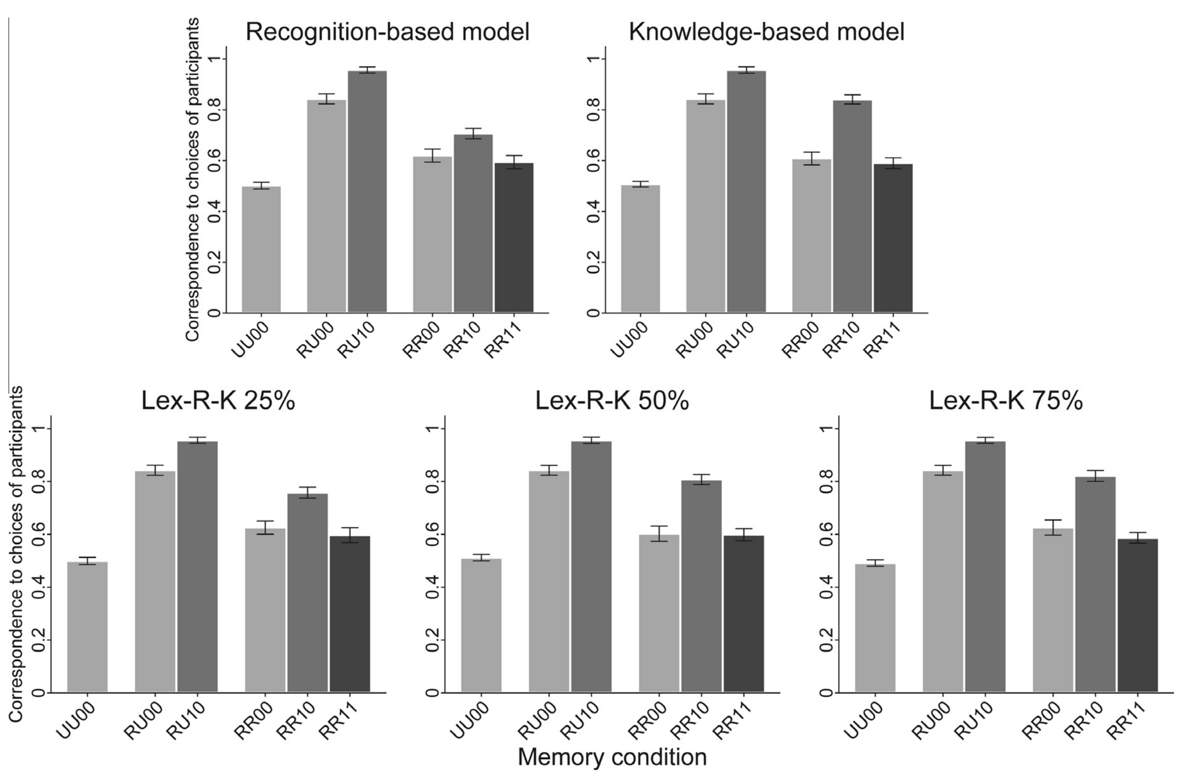

Fig. C1. Correspondence of model predictions to choices of participants. Shown are the results for the correspondence to the choices of participants (i.e., how often the models had chosen the same alternative as the participants) of the predictions of the recognition-based model (top row, left), the knowledge-based model (top row, right), and the versions of Lex-R-K that search for knowledge about 25% (bottom row, left), 50% (bottom row, middle), and 75% (bottom row, right) of the cities that each participant recognized. Bars are grouped by recognition of neither (UU), one (RU), or both (RR) cities. Darker shades of gray indicate more knowledge, that is, knowledge about neither (00), one (10), or both (11) recognized cities. Error bars indicate standard errors corrected for within-subject designs.

Table C1
Orthogonal contrasts for the differences in correspondence to the choices of participants.

Contrast	*b*	*t*	*df*	*p*	*r*
UU00 vs. RU00, RU10, RR00, RR10, RR11					
RM vs. Lex-R-K 25%, 50%, 75%, KM	0.001	1.635	616	0.103	0.07
Lex-R-K 25% vs. 50%, 75%, KM	<0.001	0.467	616	0.640	0.02
Lex-R-K 50% vs. 75%, KM	0.001	1.445	616	0.149	0.06
Lex-R-K 75% vs. KM	−0.001	−1.045	616	0.297	0.04
RU00, RU10 vs. RR00, RR10, RR11					
RM vs. Lex-R-K 25%, 50%, 75%, KM	0.001	3.369	616	0.001	0.13
Lex-R-K 25% vs. 50%, 75%, KM	0.001	1.609	616	0.108	0.06
Lex-R-K 50% vs. 75%, KM	0.001	0.868	616	0.386	0.03
Lex-R-K 75% vs. KM	<0.001	0.155	616	0.877	0.01
RU00 vs. RU10					
RM vs. Lex-R-K 25%, 50%, 75%, KM	<0.001	<0.001	616	>0.999	<0.001
Lex-R-K 25% vs. 50%, 75%, KM	<0.001	<0.001	616	>0.999	<0.001
Lex-R-K 50% vs. 75%, KM	<0.001	<0.001	616	>0.999	<0.001
Lex-R-K 75% vs. KM	<0.001	<0.001	616	>0.999	<0.001
RR00 vs. RR10, RR11					
RM vs. Lex-R-K 25%, 50%, 75%, KM	0.004	4.180	616	<0.001	0.17
Lex-R-K 25% vs. 50%, 75%, KM	0.004	3.258	616	0.001	0.13
Lex-R-K 50% vs. 75%, KM	−0.001	−0.584	616	0.560	0.02
Lex-R-K 75% vs. KM	0.005	1.751	616	0.081	0.07
RR10 vs. RR11					
RM vs. Lex-R-K 25%, 50%, 75%, KM	−0.010	−6.985	616	<0.001	0.27
Lex-R-K 25% vs. 50%, 75%, KM	−0.009	−4.677	616	<0.001	0.19
Lex-R-K 50% vs. 75%, KM	−0.006	−2.103	616	0.036	0.08
Lex-R-K 75% vs. KM	−0.004	−0.903	616	0.367	0.04

Note. This table breaks down the two-way interaction effect between memory condition (i.e., UU00, RU00, RU10, RR00, RR10, and RR11) and model (i.e., RM, Lex-R-K 25%, Lex-R-K 50%, Lex-R-K 75%, and KM) on the correspondence of the model predictions to the choices of participants. RM = recognition-based model; Lex-R-K 25%, 50%, or 75% = versions of Lex-R-K that search for knowledge about 25%, 50%, or 75% of the cities that each participant recognized; KM = knowledge-based model. UU = neither city recognized; RU = one city recognized; RR = both cities recognized; the notations 00, 10, and 11 indicate that knowledge was available for neither, one, and both cities, respectively.

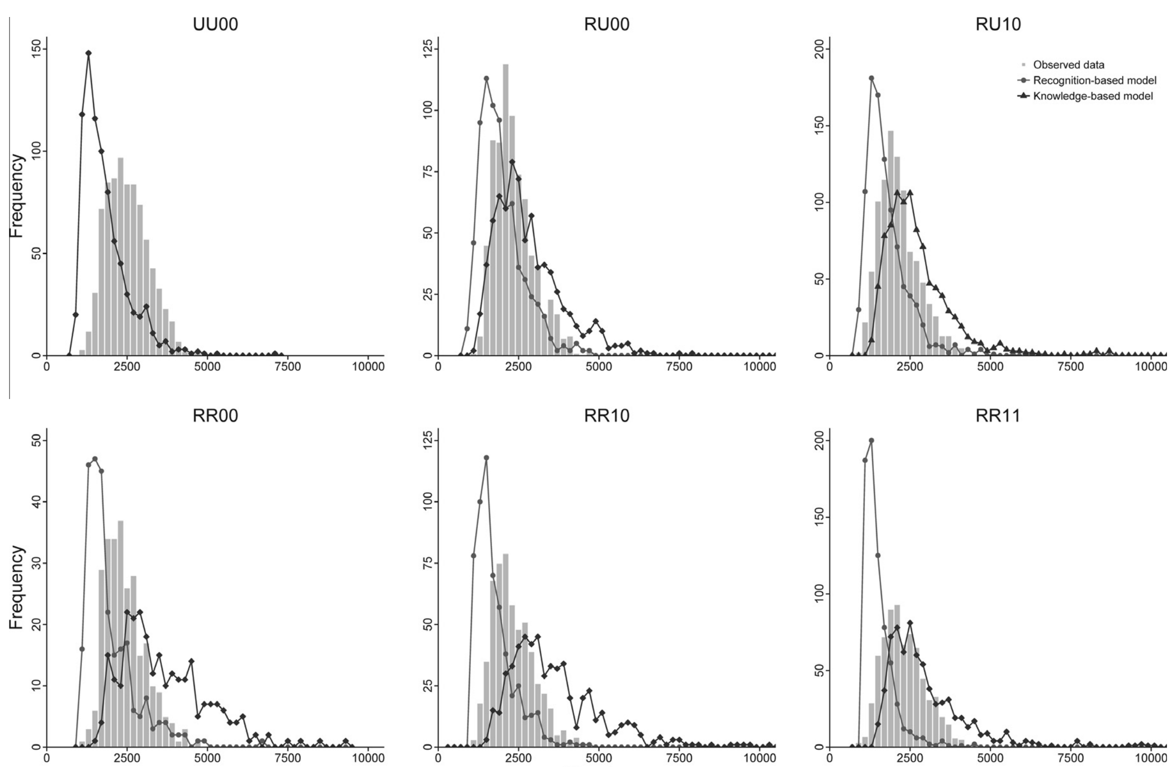

Fig. D1. Observed and predicted response time distributions for the recognition-based model and the knowledge-based model. Shown are the response time distributions of participants (gray bars), and the predictions of the recognition-based model (light gray line) and the knowledge-based model (dark gray line), computed for bin widths of 200 ms and cut off at 10 s. The graphs are split for the six memory conditions of the study, that is, whether neither (UU), one (RU), or both (RR) cities were recognized, and knowledge about neither (00), one (10), or both (11) recognized cities was available. These results for response time distributions correspond to the aggregated results shown in the top row of Fig. 3.

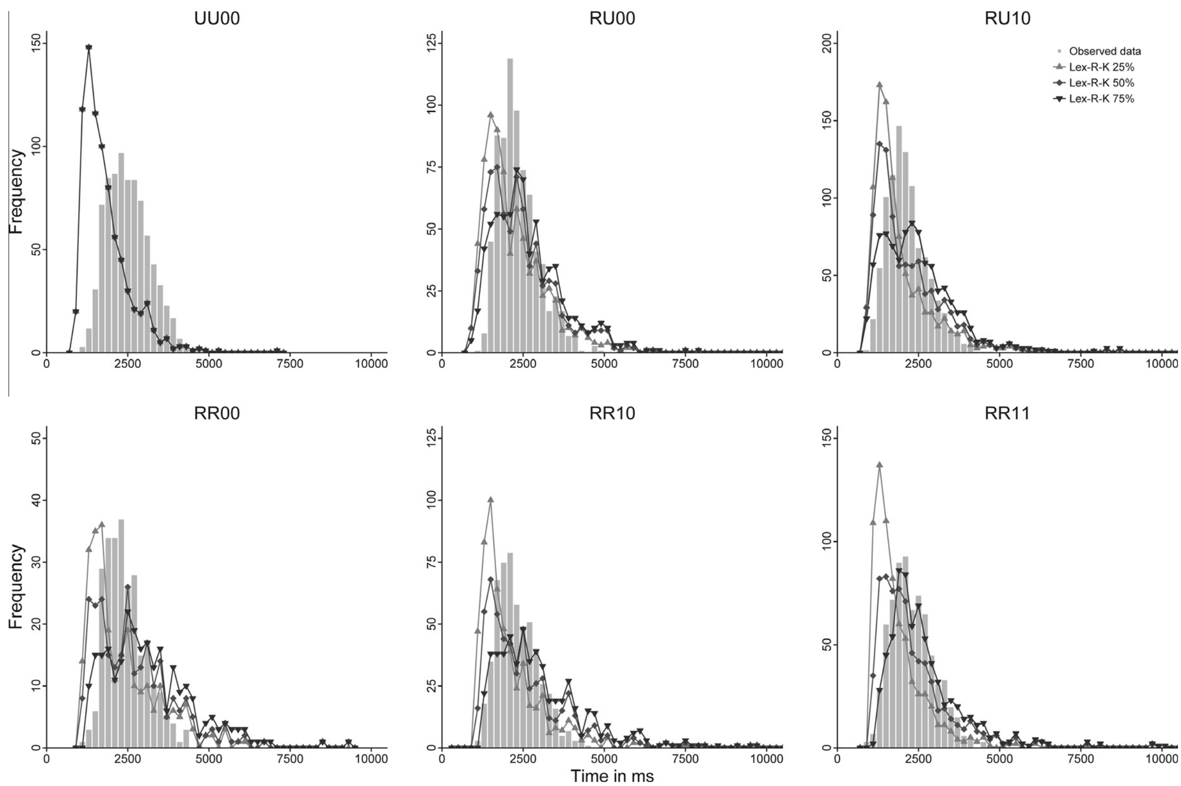

Fig. D2. Observed and predicted response time distributions for the different versions of Lex-R-K. Shown are the response time distributions of participants (gray bars) and the predictions of the versions of Lex-R-K that search for knowledge about 25% (light gray line), 50% (medium gray line), and 75% (dark gray line) of the cities that each participant recognized, computed for bin widths of 200 ms and cut off at 10 s. The graphs are split for the six memory conditions of the study, that is, whether neither (UU), one (RU), or both (RR) cities were recognized, and knowledge about neither (00), one (10), or both (11) recognized cities was available. These results for response time distributions correspond to the aggregated results shown in the bottom row of Fig. 3.

available about one compared to about both cities (i.e., RR10 vs. RR11). Similarly, the 25% version of Lex-R-K showed a lower correspondence than the models with more knowledge search when there was knowledge available about one or both cities compared to when there was no knowledge available (i.e., RR10, RR11 vs. RR00), and when there was knowledge available about one compared to about both cities (i.e., RR10 vs. RR11). In addition, the correspondence of the 50% version of Lex-R-K to choices of participants was slightly lower than the correspondence of the models with more knowledge search when there was knowledge available about one compared to about both cities (i.e., RR10 vs. RR11). This may indicate that the recognition-based model, the 25% version of Lex-R-K, and to a lesser extent also the 50% version of Lex-R-K underestimated how often participants searched for and relied on knowledge to make their decisions when they recognized both cities and could retrieve knowledge about one of them. This result is in line with the findings from the response times (which favored the three versions of Lex-R-K over the other models, with the 50% version of Lex-R-K giving the best approximation to the data) and BOLD response (which favored the 50% and 75% versions of Lex-R-K and the knowledge-based model).

Appendix D. Response time distributions in the inference task

In this Appendix, we present the response time distributions for the inference task that correspond to the aggregated response time results shown in Fig. 3. These distributions allow additional insight into why and in which memory conditions the models explain the data well and where they fail to do so.

Results for the condition in which participants recognized neither of the cities (i.e., UU00) are shown in the top left of Figs. D1 and D2. In this condition, all models only attempt to recognize both cities, and after failing to recognize them, make a choice between the cities based on guessing. With this mechanism all models systematically underestimate the observed response times. What processes might be operating in the UU00 condition that lead to the rather long response times of participants? One possibility is that response times were prolonged by repeated failures to recognize the cities. Such behavior is not captured by the ACTR models, as they do not attempt any further retrievals once one retrieval attempt has failed and may therefore predict response times that are shorter than the observed times. Further, it is possible that in the UU00 condition people search for information associated or similar to the city names in memory, or that processes unrelated to memory-based decision making are operating. Given that the exact processes in this condition were unknown and may differ between participants, we kept the mechanism implemented in the models but report the *RMSD* and *r* computed with and without the UU00 condition.

For the remaining memory conditions, where one or both cities were recognized, Fig. D1 shows that the recognition-based model systematically underestimated response times, while the knowledge-based model overestimated them. Fig. D2 illustrates that the predictions of the different versions of Lex-R-K lay between those of the recognition-based and knowledge-based models and more closely approximated the observed response time distributions than those of the other two models. These results were obtained without further scaling the retrieval times that we had estimated for the recognition and knowledge tasks; therefore they are bound to the assumption that the memory retrieval processes in the recognition and knowledge tasks give a good estimate for the processes employed for recognition and knowledge retrieval during the inference task.

Table E1
Number of participants whose response times were best accounted for by each model.

Index of model performance	Strategy model							
	Recognition-based model	Knowledge-based model	Lex-R-K model all versions	Lex-R-K model 25%	Lex-R-K model 50%	Lex-R-K model 75%	More than one model	None of the models
RMSD without UU00	4	3	13	2	8	3	7	–
RMSD with UU00	4	2	11	2	7	2	10	–
r without UU00	2	1	21	8	8	5	3	0
r with UU00	9	1	14	8	5	1	1	2

Note. Percentages for Lex-R-K indicate that the version of Lex-R-K searches for knowledge about 25%, 50%, or 75% of the cities that each participant recognized. UU00 = neither city was recognized and there was no knowledge available for either city.

Appendix E. Model performance based on individual response times

To explore how well the models accounted for the data of individual participants, we examined the model performance in terms of the *RMSD* and *r* for individual response times. Participants were classified according to which strategy model explained their response times across the memory conditions best (i.e., without UU00 and with UU00) as shown in Table E1. For the *RMSD*, participants were classified as users of a strategy when their *RMSD* value for one of the strategies was smaller than the *RMSD* values for the other strategies and the difference between the *RMSD* values was larger than a threshold that was set at 1 *SE* of the mean model fits; otherwise participants were classified such that more than one strategy model explained their data (cf. von Helversen, Mata, & Olsson, 2010). For *r*, participants were classified as users of a strategy when their *r* value for one of the strategies was larger than the *r* values for the other strategies. When two or more models explained the data equally well (i.e., their difference in *r* was smaller than 0.01), participants were classified such that more than one model explained their data. When the *r* for all models was small (i.e., smaller than 0.1), participants were classified such that none of the models explained their data.

In terms of the *RMSD*, when considering all memory conditions including the condition where neither city was recognized (i.e., UU00), for 11 of 27 participants, a version of Lex-R-K accounted best for their data (25%: 2 participants; 50%: 7 participants; 75%: 2 participants), whereas for four participants the recognition-based model and for two participants the knowledge-based model provided the best account. For the data of 10 other participants more than one model performed similarly well. When considering only memory conditions where one or both decision alternatives were recognized (i.e., a computation of the *RMSD* without UU00), for 13 of 27 participants, a version of Lex-R-K accounted best for their data (25%: 2 participants; 50%: 8 participants; 75%: 3 participants), whereas for four participants the recognition-based model and for three participants the knowledge-based model provided the best account. For the data of seven other participants more than one model performed similarly well.

In terms of *r*, when considering all memory conditions including the condition where neither city was recognized (i.e., UU00)—which led to a relatively poor account of the response time data for the models—for 14 of 27 participants, a version of Lex-R-K accounted best for their data (25%: 8 participants; 50%: 5 participants; 75%: 1 participant), whereas for nine participants the recognition-based model and for one participant the knowledge-based model provided the best account. For the data of one other participant more than one model performed similarly well; for two other participants none of the models gave a good account. However, when only considering memory conditions where one or both decision alternatives were recognized (i.e., a computation of *r* without UU00), for 21 of 27 participants, a version of Lex-R-K accounted best for their data (25%: 8 participants; 50%: 8 participants; 75%: 5 participants), whereas for two participants the recognition-based model and for one participant the knowledge-based model provided the best account. For the data of three other participants more than one model performed similarly well. Although these results are based on only the best numerical fit, they may give an indication that Lex-R-K could be a plausible strategy not only on an aggregate level, but also for many of the individual participants.

References

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York, NY: Oxford University Press.

Anderson, J. R., Albert, M. V., & Fincham, J. M. (2005). Tracing problem solving in real time: fMRI analysis of the subject-paced Tower of Hanoi. *Journal of Cognitive Neuroscience, 17*, 1261–1274. http://dx.doi.org/10.1162/0898929055002427.

Anderson, J. R., Byrne, D., Fincham, J. M., & Gunn, P. (2008). Role of prefrontal and parietal cortices in associative learning. *Cerebral Cortex, 18*, 904–914. http://dx.doi.org/10.1093/cercor/bhm123.

Anderson, J. R., Fincham, J. M., & Douglass, S. (1999). Practice and retention: A unifying analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1120–1136. http://dx.doi.org/10.1037//0278-7393.25.5.1120.

Anderson, J. R., Fincham, J. M., Qin, Y., & Stocco, A. (2008). A central circuit of the mind. *Trends in Cognitive Sciences, 12*, 136–143. http://dx.doi.org/10.1016/j.tics.2008.01.006.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*, 396–408. http://dx.doi.org/10.1111/j.1467-9280.1991.tb00174.x.

Battal, C., Fechner, H. B., Schooler, L. J., & Volz, K. G. (2012). I guess I know the answer: Neural correlates of feeling of knowing in memory-based inference [abstract]. In *Paper presented at the 42nd Annual meeting of the society for neuroscience, New Orleans, LA*. Available online in the 2012 Neuroscience Meeting Planner, Program No. 695.02.

Borst, J. P., & Anderson, J. R. (2013). Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proceedings of the National Academy of Sciences of the United States of America, 110*, 1628–1633. http://dx.doi.org/10.1073/pnas.1221572110.

Borst, J., & Anderson, J. R. (2015). Using the ACT-R cognitive architecture in combination with fMRI data. In B. Forstmann & E. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience*. New York, NY: Springer.

Borst, J. P., Nijboer, M., Taatgen, N. A., van Rijn, H., & Anderson, J. R. (2015). Using data-driven model-brain mappings to constrain formal models of cognition. *PLoS One, 10*, e0119673. http://dx.doi.org/10.1371/journal.pone.0119673.

Borst, J. P., Taatgen, N. A., Stocco, A., & van Rijn, H. (2010). The neural correlates of problem states: Testing fMRI predictions of a computational model of multitasking. *PLoS One, 5*, e12966. http://dx.doi.org/10.1371/journal.pone.0012966.

Borst, J. P., Taatgen, N. A., & van Rijn, H. (2011). Using a symbolic process model as input for model-based fMRI analysis: Locating the neural correlates of problem state replacements. *NeuroImage, 58*, 137–147. http://dx.doi.org/10.1016/j.neuroimage.2011.05.084.

Brett, M., Anton, J.-L., Valabregue, R., & Poline, J.-B. (2002). Region of interest analysis using an SPM toolbox [abstract]. In *Paper presented at the 8th international conference on functional mapping of the human brain, Sendai, Japan*. Software retrieved from: http://marsbar.sourceforge.net.

Bröder, A., & Schütz, J. (2009). Recognition ROCs are curvilinear—Or are they? On premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 587–606. http://dx.doi.org/10.1037/a0015279.

Danker, J. F., Gunn, P., & Anderson, J. R. (2008). A rational account of memory predicts left prefrontal activation during controlled retrieval. *Cerebral Cortex, 18*, 2674–2685. http://dx.doi.org/10.1093/cercor/bhn027.

Dawes, R. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34*, 571–582. http://dx.doi.org/10.1037/0003-066X.34.7.571.

Dougherty, M. R., Franco-Watkins, A. M., & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review, 115*, 199–213. http://dx.doi.org/10.1037/0033-295X.115.1.199.

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.

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London, England: Sage.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., & Penny, W. D. (2007). *Statistical parametric mapping. The analysis of functional brain images*. London, England: Academic Press.

Frosch, C. A., McCloy, R., Beaman, C. P., & Goddard, K. (2015). Time to decide? Simplicity and congruity in comparative judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*, 42–54. http://dx.doi.org/10.1037/a0037411.

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

Gigerenzer, G., & Goldstein, D. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making, 6*, 100–121.

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

Glöckner, A., Hilbig, B. E., & Jeckel, M. (2014). What is adaptive about adaptive decision making? A parallel constraint satisfaction account. *Cognition, 133*, 641–666. http://dx.doi.org/10.1016/j.cognition.2014.08.017.

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

Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer, P. M. Todd, & The ABC research group (Eds.), *Simple heuristics that make us smart* (pp. 37–58). New York, NY: Oxford University Press.

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. http://dx.doi.org/10.1037/a0013025.

Herzog, S., & Hertwig, R. (2013). The ecological validity of fluency. In C. Unkelbach & R. Greifeneder (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behavior* (pp. 190–219). New York, NY: Psychology Press.

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. http://dx.doi.org/10.1037/a0017518.

Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2011). Fluent, fast, and frugal? A formal model evaluation of the interplay between memory, fluency, and comparative judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 827–839. http://dx.doi.org/10.1037/a0022638.

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. http://dx.doi.org/10.1037/a0016565.

Horn, S. S., Ruggeri, A., & Pachur, T. (in press). The development of adaptive decision making: Recognition-based inference in children and adolescents. *Developmental Psychology., 132*, 133–156 (in press), http://dx.doi.org/10.1037/dev0000181.

Juslin, P., Olsson, H., & Olsson, A.-C. (2003). Exemplar effects in categorization and multiple-cue judgment. *Journal of Experimental Psychology: General, 132*, 133–156. http://dx.doi.org/10.1037/0096-3445.132.1.133.

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

Luan, S., Schooler, L. J., & Gigerenzer, G. (2014). From perception to preference to inference: An approach-avoidance analysis of thresholds. *Psychological Review, 121*, 501–525. http://dx.doi.org/10.1037/a0037025.

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 and Review, 17*, 287–309. http://dx.doi.org/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.

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

McCloy, R., Beaman, C. P., Frosch, C. A., & Goddard, K. (2010). Fast and frugal framing effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1043–1052. http://dx.doi.org/10.1037/a0019693.

Morey, R. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology, 4*, 61–64.

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. http://dx.doi.org/10.1037/0278-7393.30.4.923.

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

Pachur, T., & Biele, G. (2007). Forecasting from ignorance: The use and usefulness of recognition in lay predictions of sports events. *Acta Psychologica, 125*, 99–116. http://dx.doi.org/10.1016/j.actpsy.2006.07.002.

Pachur, T., Bröder, A., & Marewski, J. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making, 21*, 183–210. http://dx.doi.org/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. http://dx.doi.org/10.1037/0278-7393.32.5.983.

Pachur, T., Mata, R., & Schooler, L. J. (2009). Cognitive aging and the adaptive use of recognition in decision making. *Psychology and Aging, 24*, 901–915. http://dx.doi.org/10.1037/a0017211.

Pachur, T., & Olsson, H. (2012). Type of learning task impacts performance and strategy selection in decision making. *Cognitive Psychology, 65*, 207–240. http://dx.doi.org/10.1016/j.cogpsych.2012.03.003.

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 Psychology, 2*, 147. http://dx.doi.org/10.3389/fpsyg.2011.00147.

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

Pinheiro, J., Bates, D., DebRoy, S., & Sarkar, D. The R Development Core Team. (2012). nlme: Linear and nonlinear mixed effects models (R package version 3.1-104) Retrieved from: http://cran.r-project.org/web/packages/nlme/.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review, 109*, 472–491. http://dx.doi.org/10.1037/0033-295X.109.3.472.

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. http://dx.doi.org/10.1037/a0019737.

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

R Development Core Team (2012). *R: A language and environment for statistical computing (version 2.15.1)*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from: http://www.r-project.org/.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134. http://dx.doi.org/10.1037/0033-295X.88.2.93.

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

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. http://dx.doi.org/10.1177/0956797611417454.

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

Schwikert, S. R., & Curran, T. (2014). Familiarity and recollection in heuristic decision making. *Journal of Experimental Psychology: General, 6*, 2341–2365. http://dx.doi.org/10.1037/xge0000024.

Sohn, M.-H., Goode, A., Stenger, V. A., Carter, C. S., & Anderson, J. R. (2003). Competition and representation during memory retrieval: Roles of the prefrontal cortex and the posterior parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 100*, 7412–7417. http://dx.doi.org/10.1073/pnas.0832374100.

Sohn, M.-H., Goode, A., Stenger, V. A., Jung, K.-J., Carter, C. S., & Anderson, J. R. (2005). An information-processing model of three cortical regions: Evidence in episodic memory retrieval. *NeuroImage, 25*, 21–33. http://dx.doi.org/10.1016/j.neuroimage.2004.11.001.

Stocco, A., & Anderson, J. R. (2008). Endogenous control and task representation: An fMRI study in algebraic problem solving. *Journal of Cognitive Neuroscience, 20*, 1300–1314. http://dx.doi.org/10.1162/jocn.2008.20089.

Taatgen, N. A., van Rijn, H., & Anderson, J. R. (2007). An integrated theory of prospective time interval estimation: The role of cognition, attention, and learning. *Psychological Review, 114*, 577–598. http://dx.doi.org/10.1037/0033-295X.114.3.577.

Turner, B. M., Forstmann, B. U., Love, B. C., Palmeri, T. J., & van Maanen, L. (in press). Approaches to analysis in model-based cognitive neuroscience. *Journal of Mathematical Psychology, 2*, 7–8 (in press), http://dx.doi.org/10.1016/j.jmp.2016.01.001.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency. *Cognitive Psychology, 5*, 207–232. http://dx.doi.org/10.1016/0010-0285(73)90033-9.

Van Zandt, T. (2000). ROC curves and confidence judgements in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 582–600. http://dx.doi.org/10.1037/0278-7393.26.3.582.

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. http://dx.doi.org/10.1162/jocn.2006.18.11.1924.

von Helversen, B., Mata, R., & Olsson, H. (2010). Do children profit from looking beyond looks? From similarity-based to cue abstraction processes in multiple-cue judgment. *Developmental Psychology, 46*, 220–229. http://dx.doi.org/10.1037/a0016690.

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1235–1253. http://dx.doi.org/10.1037/0278-7393.19.6.1235.

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis (R package version 0.9.3.1)*. New York, NY: Springer. Retrieved from: http://ggplot2.org.