CrossMark

Use of the familiarity difference cue in inferential judgments

Ping Xu[1] · Claudia González-Vallejo[1] · Justin Weinhardt[2] · Janna Chimeli[1] ·
Figen Karadogan[3]

Abstract The familiarity difference cue has been regarded as a general cue for making inferential judgments (Honda, Abe, Matsuks, & Yamagishi in *Memory and Cognition*, *39*(5), 851–863, 2011; Schwikert & Curran in *Journal of Experimental Psychology: General, 143*(6), 2341–2365, 2014). The current study tests a model of inference based on familiarity differences that encompasses the recognition heuristic (Goldstein & Gigerenzer, 1999, Goldstein & Gigerenzer in *Psychological Review, 109*(1), 75–90, 2002). In two studies, using a large pool of stimuli, participants rated their familiarity of cities and made choices on a typical city-size task. The data were fitted with the *r-s* model (Hilbig, Erdfelder, & Pohl in, *Journal of Experimental Psychology: Learning Memory and Cognition, 37*(4), 827–839, 2011), which was adapted to include familiarity differences. The results indicated that people used the familiarity difference cue because the participants ignored further knowledge in a substantial number of cases when the familiarity difference cue was available. An analysis of reaction-time data further indicated that the response times were shorter for heuristic judgments than for knowledge-only-based judgments. Furthermore, when knowledge was available, the response times were shorter when knowledge was congruent with a heuristic cue than when it was in conflict with it. Differences between the familiarity difference cue and

the fluency heuristic (Schooler & Hertwig, 2005, *Psychological Review, 112*, 610–628) are discussed.

Comparative judgments are ubiquitous in many decision-making domains. If concerned with diet, then one may ask: Which dish has more calories? A swordfish dinner with Asian style spinach and sticky rice, or a seafood risotto? A decision-maker could, in principle, calculate the calories with known ingredients and amounts, but, most likely, the decision will follow an inference. Similarly, answering a general knowledge question such as, "Which city has a larger population: (a) Heidelberg or (b) Bonn?" (Gigerenzer, Hoffrage, & Kleinbolting, 1991) requires bringing relevant information to mind and making an inference because the exact answer is unknown to the decision-maker. The recognition heuristic (RH) is a fast and frugal rule proposed as a key mechanism for making comparative judgments in knowledge domains such as this city-size task (Goldstein & Gigerenzer, 1999, 2002). The rule simply states that individuals select the recognized city without considering further information.

The RH is assumed to work well because of its high correlation with the to-be-judged criterion in a natural environment (i.e., the actual population in the city-size task). For example, recognized cities are usually more populous compared to unrecognized cities, hence selecting a recognized city over an unrecognized city can lead to correct judgment in a city-size task. The relative frequency with which a cue correctly predicts a target is termed the "cue validity" (Goldstein & Gigerenzer, 1996). Many studies have tested the RH, and support for it varies depending on the characteristics of the

✉ Ping Xu
px483811@ohio.edu

[1] Department of Psychology, Ohio University, Athens, OH 45701, USA

[2] Haskayne School of Business, University of Calgary, Calgary, AB, Canada

[3] Department of Psychology, Central Michigan University, Mt. Pleasant, MI, USA

task (Gigerenzer & Goldstein, 2011a, 2011b; Goldstein & Gigerenzer, 1999, 2002; Oppenheimer, 2003; Pachur, Todd, Gigerenzer, Schooler, & Goldstein, 2011).

A main criticism of the RH stems from its assumption that a participant using the RH ignores other information on the recognized object even if it is available (Goldstein & Gigerenzer, 1999, 2002). The rate of using the RH is frequently measured by the adherence ratio, which is the proportion of inferences that are in line with selecting the recognized object (Goldstein & Gigerenzer, 1999). Researchers have pointed out, however, that considering knowledge may also lead to the choice of the recognized object (Hilbig, Erdfelder, & Pohl, 2010), hence the adherence ratio may not be a valid measure of the rate of using the recognition cue in isolation. To provide a less biased measure of RH use, a multinomial processing tree model was developed by Hilbig et al. (2010). The model was further extended to incorporate unbiased measurements of other heuristics (Castela, Kellen, Erdfelder, & Hilbig, 2014; Hilbig, Erdfelder, & Pohl, 2011).

Another limitation of the RH is that it applies when only one object is recognized. When both objects are recognized, people need to rely on other strategies. One strategy suggested by researchers is the fluency heuristic (FH; Schooler & Hertwig, 2005). The FH states that people can compare retrieval times for the two objects being compared and select the object with faster recognition time. Hertwig and his colleagues (Hertwig, Herzog, Schooler, & Reimer, 2008) showed that people's inferences were in line with the FH as participants were more likely to choose an object that was recognized faster. In addition, the cue validity for the FH was higher than .55 in three different domains, implying that the correct judgment rate is higher than chance. However, using a multinomial processing tree model (the *r-s* model), Hilbig et al. (2011) found participants used the FH in isolation only 23% of the time. Hilbig et al.'s individual data analysis revealed that the majority of participants never used the FH in isolation. Two studies conducted by Schwikert and Curran (2014) obtained similar low rates of using the FH.

A related construct to both the RH and FH is the level of subjective familiarity that individuals have for items. Whittlesea (1993) stated that a feeling of familiarity is a mental resonance of prior events when encountering a recurrent event. An activation of memory trace related to the stimuli is a precondition for a feeling of familiarity. In the recognition memory literature, familiarity has been regarded as a memory process that contributes to the recognition process (Park & Yonelinas, 2007; Yonelinas, 1994). The process can be described by a signal detection model as proposed by Yonelinas (1994). More specifically, in a typical recognition task, people usually make recognition judgments among a mixture of studied and new items. Familiarity is assumed to vary from item to item, so the subjective familiarity of items is continuous and follows a normal distribution. The normal distributions of the new and old items are different but likely overlap, such that when people make judgments according to a specific familiarity criterion, misses and false alarms may occur (Yonelinas, 1994). Congruent with this model, familiarity is regarded as the underlying mechanism of the RH (Gigerenzer et al., 1991; Schwikert & Curran, 2014).

The relationship between a feeling of familiarity and the concept of fluency has not been clearly addressed in the literature. This may be in part because the concept of "fluency" follows different interpretations across studies. Oppenheimer (2008) interpreted it as "the subjective experience of ease" (p. 237). According to this definition, there can be many different types of fluency (perceptual, conceptual, linguistic, retrieval, decision; Oppenheimer, 2008; for a review, also see Alter & Oppenheimer, 2009). Hence, fluency is a broad concept. In addition, fluency may be a factor that generates feelings of familiarity. Studies have shown that types of fluency, such as perceptual and conceptual fluency, lead to feelings of familiarity even in the absence of prior experience (Whittlesea, 1993; Whittlesea, Jacoby, & Girard, 1990). On the other hand, Schooler and Hertwig's (2005) work focused specifically on retrieval fluency, which is defined as the retrieval time when identifying an object. In their interpretation, a feeling of familiarity is associated with retrieval fluency, which is attributed to prior experience. In other words, retrieval time is a proxy for a feeling of familiarity. This interpretation implies that familiarity and fluency are highly associated.

Starting from these notions, Schwikert and Curran (2014) studied the memory processes that contribute to the RH and the FH. Schwikert and Curran (2014) first recorded electroencephalogram (EEG) data while participants performed a timed recognition and a population inference task, and they found that the RH was used in the majority of cases while the FH was used in only 16% of cases, as estimated by the *r-s* model. In addition, they found different event-related potentials (ERP) for the RH-based and the FH-based decisions. Familiarity appeared to play an important role in the RH-based but not in the FH-based decisions; the latter relying more on recollection knowledge. In their second experiment, Schwikert and Curran categorized cities as unrecognized (U): participants had not heard of the city; familiar (F): participants had heard of the city but did not have further knowledge; and remembered (R): participants had heard of the city and were able to provide additional information about the city. This latter type was conceived as having greater familiarity than the F type. The authors again found that the RH was used in isolation in the majority of cases, whereas the FH was used only in about 21% of the cases. Additionally, the authors found that participants used the RH in a greater proportion of remembered–unrecognized (RU) pairs than of familiar–unrecognized (FU) pairs. They also found that participants used the FH more in remembered–familiar (RF) pairs (28%) than in remembered–remembered (RR; 17%) and familiar–familiar

(FF; 16%) pairs. In other words, they found that both the RH and the FH were more likely in places where the difference of items' familiarity was larger (i.e., for RU and RF pairs).

Based on their findings, Schwikert and Curran (2014) advanced a new theoretical framework for memory based on familiarity differences. As shown in Fig. 1, Schwikert and Curran's model assumes that when the familiarity difference is large between two items, people choose the more familiar item without any consideration of other recollected knowledge; when the familiarity difference is low, depending on whether knowledge is available, people rely on either knowledge or a retrieval fluency difference to make the inference. Guessing occurs only when none of the above strategies are available.

The notion of a familiarity difference echoes that of the memory-state heuristic, which states that people make decisions by choosing the object with a stronger memory state (Erdfelder, Küpper-Tetzel, & Mattern, 2011). The assumption of this heuristic is that memory strength is usually highly correlated with a to-be-judged criterion, and the reliance on it increases with the memory strength distance between the two objects. Castela et al. (2014) tested this hypothesis. They found that people were more likely to use a recognition heuristic when the recognized object was recognized with further knowledge (i.e., an object with a stronger memory strength) than when it was merely recognized (i.e., an object with a lower memory strength). In addition, they found that when making choices between a highly recognized object and a merely recognized object, people sometimes selected the

highly recognized object without incorporating other knowledge. In other words, their study showed that memory strength differences can be used as cues in making inferences when both objects are recognized.

Honda, Abe, Matsuks, and Yamagishi (2011) studied the idea that familiarity difference is a more general cue than the RH and the FH. Congruent with Schwikert and Curran's (2014) proposition, Honda et al. argued that when both objects are recognized, people choose the more familiar object from a pair because that object is likely to have a higher value on the to-be-judged criterion (i.e., the actual population in the city-size task). In their study, participants rated their knowledge about cities in an ordinal scale and then made a series of binary choices in a standard city-size task. The results revealed that the familiarity difference, as measured by knowledge ratings of the cities, was highly correlated with the proportion of subjects who chose the larger city. Regression-model analysis revealed that the familiarity difference was a critical predictor of the choice for both the recognized–unrecognized pairs and the recognized–recognized pairs. Their model comparison showed that a knowledge-based model outperformed the familiarity-based model when both objects were recognized. This implies that both familiarity difference and further knowledge may play a role in the final decision. Other researchers have also presented evidence that individuals use additional information other than a heuristic strategy when making selections (Hilbig & Pohl, 2008; Hilbig, Pohl, & Broder, 2009; Newell & Fernandez, 2006; Newell & Shanks, 2004; Oppenheimer, 2003; Pohl, 2006; Pachur &
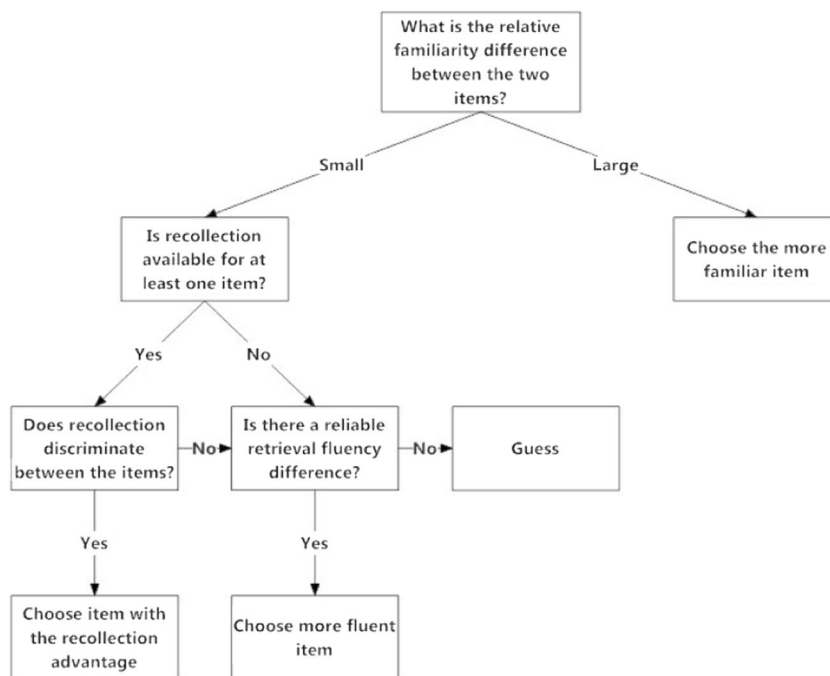

Fig. 1 The memory-based decision flow chart. Reprinted from "Familiarity and Recollection in Heuristic Decision Making," by Schwikert and Curran, 2014, *Journal of Experimental Psychology: General*, p. 2344. Copyright 2014 by American Psychologist Association

Hertwig, 2006; Richter & Spath, 2006; Schwikert & Curran, 2014).

The *r-s* model

The *r-s* model is a multinomial-tree model of the judgment and choice process of responses to tasks like the city-size task. This study use the *r-s* model to test the use of familiarity differences. We first introduce the model as developed by Hilbig et al. (2011), then discuss its application to the current study.

The *r-s* model is comprised of four trees (see Fig. 2), representing four recognition states of a comparative judgment: (a) Both items are recognized, and the difference of their retrieval times is below a threshold; (b) both items are recognized, and the difference of their retrieval times is above a threshold; (c) one item is recognized; (d) neither item is recognized (Hilbig et al., 2011). When both items are recognized, they are further divided into fluency-homogeneous cases and fluency-heterogeneous cases, as shown on the first two nodes in Fig. 2, and the division is based on a threshold. Both studies, Hilbig et al. (2011) and Schwikert and Curran (2014), used a threshold of 100 milliseconds (ms). That is, cases with retrieval difference above 100 ms were classified as fluency-heterogeneous cases, whereas those below 100 ms were fluency-homogeneous cases. This classification is in accordance with the finding that participants do not detect retrieval differences below 100 ms (Hertwig et al., 2008).

The model further assumes that for fluency-homogeneous cases people use knowledge to make inferences, with valid knowledge leading to correct judgments and invalid knowledge leading to false judgments. The processing logic for the fluency-heterogeneous tree is the same as that of the recognition tree. Specifically, for cases classified into these two trees, people can either use the RH/FH only, or they can incorporate knowledge to make inferences. The former means that people adhere to the RH/FH, and make a correct judgment if the RH/FH is valid. For the latter, people make correct judgments when the knowledge is valid. Note that even when people use knowledge to make decisions, their decisions can be either in line with the RH/FH or not, depending on whether the RH/FH are valid or not. For cases in which neither item is recognized, people rely on guessing to make judgments, with valid guesses leading to correct choices.

Taken together, the model comprises 12 observable outcomes determined from nine parameters, as shown in Fig. 2. The *b1* parameter is a knowledge-validity parameter for items with equal recognition levels. The *c* parameter indicates the validity of the FH, and the *b2* parameter indicates the knowledge validity within the cases for which the FH is available as a cue. The *a* parameter is the validity of the RH. The knowledge-validity parameter of the recognition tree (*b3*)

was assumed to be an average of *b1* and *b2* by Hilbig et al. (2011). More precisely, it was defined as, $b3 = p \times b1 + (1 - p) \times b2$, where *p* is the proportion of fluency-homogeneous cases out of all knowledge cases (i.e., both fluency-homogeneous and fluency-heterogeneous cases; see Hilbig et al., 2011, p. 838). Parameters *r* and *s* represent the rate of using the RH and the FH, respectively. The parameter *g* is the rate of making correct selections via guessing. All parameters, except *b3*, are free and must be estimated to fit the model.

The current study

As stated earlier, the construct of familiarity has had a central role in many theories of memory-based judgments. Fluency has been conceived as a factor that generates feelings of familiarity (Whittlesea, 1993; Whittlesea et al., 1990). Schooler and Hertwig (2005) further interpreted feelings of familiarity as associated with retrieval fluency. That is, retrieval fluency in the *r-s* model determines the classification of items into the fluency-homogenous and fluency-heterogeneous cases. In the present study, we focus on feelings of familiarity directly, and test the model in Fig. 2 using self-reported familiarity. This measure of familiarity is worth exploring as a test of the generalizability of the findings of the *r-s* model based on fluency. In addition, we explore the role of familiarity in making fast-and-frugal judgments similar to those made with the RH.

In previous studies, the construct of familiarity has been measured in different ways. In the study conducted by Castela et al. (2014), familiarity was not directly measured, but rather items that subjects claimed to have further knowledge of were classified as more familiar than items without that knowledge. In the study conducted by Schwikert and Curran (2014), familiarity was measured by retrieval time or by categorizing items into recognized, familiar, and remembered items with the assumption that *remembered* items were more familiar than the *familiar* items. In the study conducted by Honda et al. (2011), familiarity was reported by participants using a 5-point Likert-rating scale. We use the Honda et al. measurement and assume a continuous familiarity variable, which is congruent with the notion used by memory researchers (e.g., Park & Yonelinas, 2007; Yonelinas, 1994).

We note that self-reports are often a poor proxy of inner psychological states, such as measures of self-knowledge or metacognition (e.g., overconfidence phenomenon, Fischhoff, Slovic, & Lichtenstein 1977; and the KrugerDunning effect, Kruger & Dunning, 1999). On the other hand, self-reports can be better measures of inner states as demonstrated by Reid and González-Vallejo (2009) in risky decision-making. Compared to physiological responses, self-reported measures of attribute importance were better predictors of individuals' choices. Thus, the value of self-reports in some domains provides support for assessing their role in inferential judgments. We
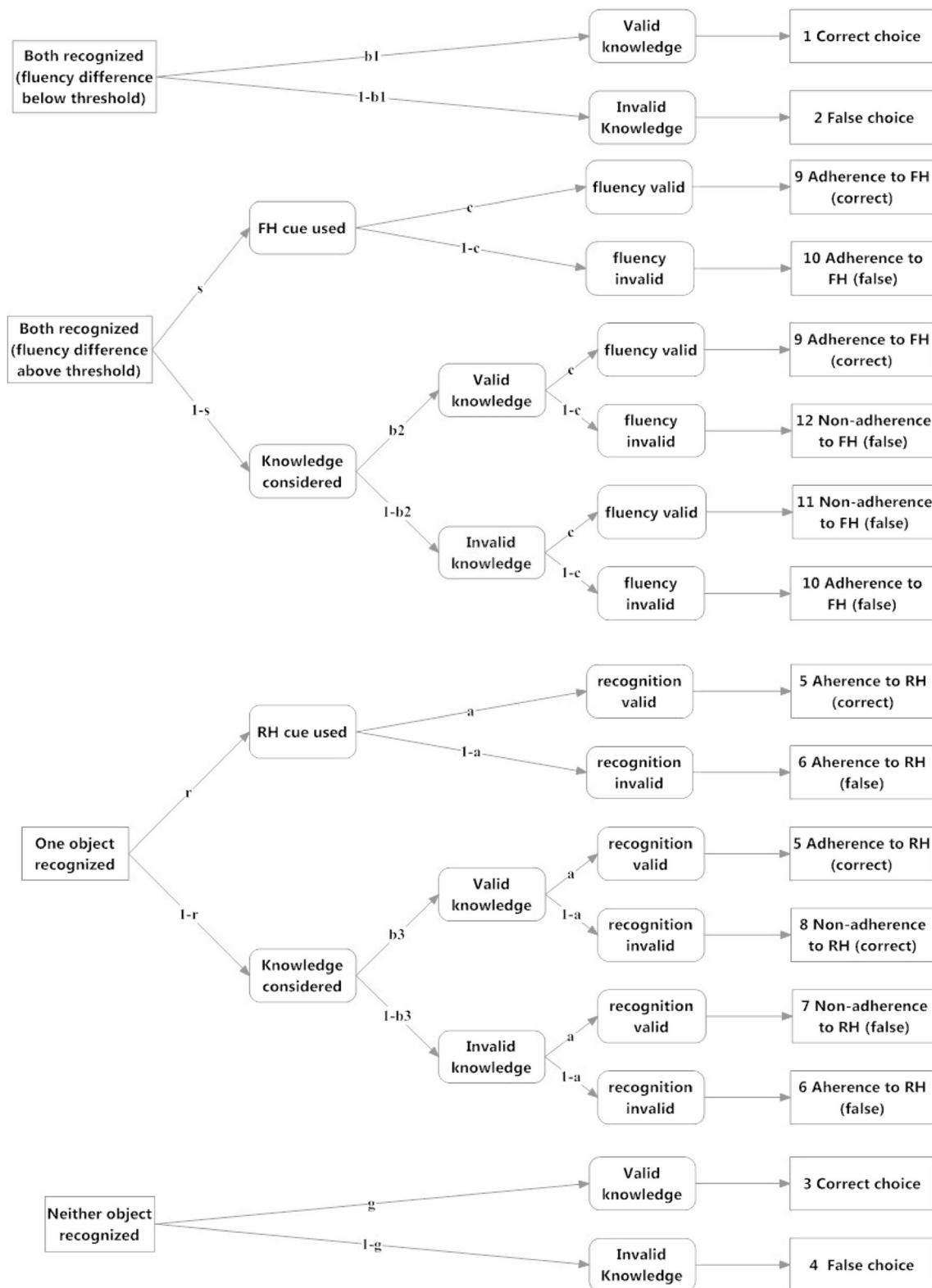
Fig. 2 Processing trees representing the *r-s* model. Parameters include recognition validity (*a*), fluency validity (*c*), knowledge validity (*b1*, *b2*, *b3*), probability of valid guesses (*g*), probability of using the fluency heuristic (FH; *s*), probability of using the recognition heuristic (RH; *r*). Boxes with rounded corners signify latent states. Processing tree reprinted from "Fluent, Fast, and Frugal? A Formal Model Evaluation of the Interplay Between Memory, Fluency, and Comparative Judgments," by Hilbig, Erdfelder, and Pohl, 2011, *Journal of Experimental Psychology: Learning, Memory, and Cognition,* p. 830. Copyright 2011 by the American Psychological Association

investigate their validity in predicting choices in the city-size task.

The *r-s* model in tests of familiarity differences

In the current study, we use the *r-s* model to test the use of familiarity differences in situations in which both items are recognized. That is, in the adaptation of the *r-s* model to our situation, the classification of homogeneity is not based on fluency but rather on the direct measure of familiarity differences. Furthermore, we investigate whether familiarity difference, as a cue, corresponds with a fast-and-frugal process that leads to inferences without further consideration of knowledge.

One method for testing whether a heuristic is fast and frugal is to use response times relying on the assumption that fast responses are evidence of little thinking. Pachur and Hertwig (2006) found that response times were shorter when a comparative judgment was in line with the RH than when it was not, supporting the idea that the RH is a fast-and frugal heuristic. However, response time is not a pure processing measure. Several studies found that within recognition cases, people make quicker judgments when they have more knowledge about the recognized item, or are more familiar with the recognized item (Hilbig & Polh, 2009; Richter & Spath, 2006; Schweickart & Brown 2014). This finding implies that using knowledge, in addition to the RH, may also be a fast, albeit not frugal, process. Furthermore, Glöckner and Betsch (2012) found that adding information can result in faster decisions if the added information is in line with other information. Therefore, Experiment 2 tests the fast nature of familiarity differences by relying on the *r-s* model parameters and processing assumptions in addition to response time. Model details in the Method section will make this clear.

Our adaptation of the *r-s* model in Fig. 2 uses the following pair classification. The familiarity-homogeneous tree (the first tree) includes: F_F (both objects recognized with low familiarity ratings) and F+_F+ pairs (both objects recognized with higher familiarity ratings). The familiarity-heterogeneous tree (the second tree) includes: F_F+ (both objects recognized but with different levels of familiarity ratings). The recognition tree (the third tree) includes: F_U (one recognized with a low familiarity rating and the other is unrecognized) and F+_U (one recognized with a high familiarity rating and the other is unrecognized). The guessing tree (the fourth tree) includes: U_U (both unrecognized).

Following the model in Fig. 2 as adapted for this study, the parameter *s* stands for the ratio of using the familiarity difference cue in isolation, or as a simple heuristic. The parameter *r* stands for the ratio of using the RH in isolation. We use the *r-s* model to estimate *r* and *s* parameters in two experiments. Both experiments used a set of 24 large and 24 small cities in the state of Ohio, and 24 fictitious cities (see materials in the

Methods section for further details). These cities formed six pair types based on population size that allowed for estimating the key model parameters. The pair types were large_large (LL), small_small (SS), fictitious_fictitious (FF), large_small (LS), large_fictitious (LF), and small_fictitious (SF). We refer to L cities as those belonging to the large city pool of items (with an average population equal to 128,857 inhabitants) and to the S cities as those belonging to the small city pool (with an average population equal to 17,411 inhabitants). First, participants completed a familiarity rating for each city using a Likert-type scale, with greater numbers implying greater familiarity. Following the familiarity ratings, the participants selected the more populous city as a comparative judgment task.

This design allowed for testing the model with different levels of cue validity within the same domain. In addition, this design allowed for the assessment of participants' familiarity differences as well as the validity of such differences within a real environment known to the participants. The studies were also meant to explore false recognition (and false feelings of familiarity), thus the inclusion of the fictitious items. Experiment 1 provided the initial test for the model under different cue validity conditions (i.e., large, small, and fictitious city combinations). Experiment 2 was a generalization test, with the addition of a response-time analysis that explored situations in which the familiarity difference leads to fast responding with or without usage of knowledge. Specific timing predictions follow from the model event classification.

Experiment 1

Method

Materials

Using the 2000 census, the 24 largest cities from the state of Ohio (USA) were selected, with populations ranging from 37,661 (Lima) to 769,360 (Columbus). Similarly, a set of 24 small cities were selected, with populations ranging from 12,851 (Bedford) to 21,001 (Wadsworth). Finally, a set of 24 fictitious city names were created from a list of surnames (e.g., Hammond) common in Ohio. We include fictitious cities to study false recognition (e.g., judging a fictitious city as a recognized city; as in Newell & Fernandez, 2006; Oppenheimer, 2003). Using the 72 city names twice,[1] the experiment had 72 pairs classified into six pair types. There were 12 pairs of large cities (LL), 12 pairs of small cities (SS),

---

[1] In contrast to previous studies in which limited stimuli were used (Richter & Spath, 2006), meaning they repeated city names multiple times, the current study had a large set of stimuli, which is why each city name appeared only twice. This design avoids the possibility of participants becoming more and more familiar with the city names as cities repeat in presentation.

12 pairs of fictitious cities (FF), 12 pairs of large and small cities (LS), 12 pairs of large and fictitious cities (LF), and 12 pairs of small and fictitious cities (SF). The experiment had two different sets of materials, Form 1 and Form 2, created based on random pairings.

Participants

One hundred and thirteen undergraduates at Ohio University participated for course credit.

Procedure

Forty-five participants used Form 1, and others used Form 2, with random assignment to each. Participants were tested in small groups of 10 to 15 students who used paper and pencil. Booklets containing two familiarity-rating tasks and one choice task were distributed. In the familiarity-rating tasks, participants rated each of the 72 city names for feelings of familiarity using a 5-point Likert-type rating scale, with the values from 1 to 5, indicating *very unfamiliar*, *unfamiliar*, *somewhat unfamiliar*, *familiar*, and *very familiar*, respectively. In the choice task, pairs of city names appeared in horizontal lines with a hyphen between them. For each pair, participants had to circle the city that they thought had a larger population. The left or right position for larger and small cities was randomized.

The order of the tasks was: familiarity-rating task, the choice task, and a second familiarity-rating task. The second familiarity-rating task was included to check whether participants' familiarity ratings changed as a result of presentations of the cities in the choice task. There was no time limit for each task; the researcher always waited until all participants had finished one task before advancing to the next stage of the experiment.

Results

Classification of the city types

The means of the first and second familiarity ratings combined across cities were not significantly different. We thus used the first ratings in the following analyses as they are free from any influence of the exposure to names in the choice task.[2] Descriptive statistics of familiarity judgments for each type of city is shown in Table 1. Mean familiarity rating for the large cities is higher than that of the small cities, which is higher than that of the fictitious cities, as was expected. The

---

[2] Although a previous study showed that the use of the fluency heuristic was independent from the order of recognition and choice tasks (Pohl, Erdfelder, Michalkiewicz, Castela, & Hilbig, 2016), we used the first ratings only. This was to ensure that prior exposure to the items had no influence on choice (in line with exposure effects; e.g., Zajonc, 1968).

Table 1 Descriptive statistics of familiarity ratings for Experiment 1 and Experiment 2

	Experiment 1 Mean	SD	Experiment 2 Mean	SD
Fictitious city	1.42	.13	4.81	2.39
Small city	2.18	.45	20.91	9.78
Large city	3.28	.81	44.19	18.78

familiarity ratings of the cities across participants showed that 50.1% of cases were rated as 1, which represented very unfamiliar. For purposes of testing the r-s model, cities with familiarity rating of 1 were classified as unrecognized (U type); those with familiarity ratings equal to 2 and 3 were classified as less familiar (F type) and cities with familiarity ratings of 4 or greater were classified as more familiar (F+ type).

Correlation between city size and familiarity rating

To test whether the more familiar cities are more populous than the less familiar cities, we conducted a correlation analysis. For each of the 48 real cities, we calculated the mean familiarity rating across all 113 participants (excluding the fictitious cities). The size of the city was log transformed. There was a large and significant positive correlation between city sizes and familiarity ratings, $r(46) = .84$, $p < .001$.

We also tested the correlation at the individual level. One hundred and eleven out of 113 participants (98.23%) showed a significant positive correlation between city sizes and familiarity ratings. Hence, the correlational analyses indicated that people are more familiar with cities that have larger population sizes. In other words, the familiarity difference cue possesses a high cue validity.

Correct responding

Performance varied as a function of both familiarity and population size pair, as shown in Table 2. A chi-square test of independence was significant, $\chi^2(20) = 1206.2$, $p < .001$. Across familiarity judgment pair types, participants performed best for the LF and LS population-size types, and worst for the SS type. That is, performance was better when a city was large. Interestingly, performance for pairs including a fictitious city was good (64% and 76% correct for the SF and LF, respectively). This implies that selections of fictitious cities were not frequent, but neither were they negligible (i.e., selection of the fictitious city occurred in 36% and 24% of the cases). Across population-size types, participants performed best for the F+_U type, followed by the F+_F type. That is, pairs of cities in which one was highly recognized led to better performance.

Table 2 Frequencies of correct response of familiarity judgment type by population-size type

	SF	LF	SS	LS	LL	Total
F+_F+	14 (.64 [a])	24 (.80)	49 (.48)	159 (.73)	323 (.75)	569 (.71)
F_F	68 (.57)	43 (.61)	67 (.56)	63 (.68)	60 (.66)	301 (.61)
F+_F	59 (.69)	161 (.86)	90 (.49)	244 (.82)	216 (.73)	770 (.73)
F+_U	192 (.87)	408 (.93)	103 (.50)	297 (.89)	183 (.76)	1,183 (.82)
F_U	201 (.67)	223 (.73)	152 (.49)	151 (.71)	87 (.54)	814 (.63)
U_U	324 (.54)	170 (.53)	238 (.55)	117 (.59)	69 (.51)	918 (.54)
Total	858 (.64)	1,029 (.76)	699 (.52)	1,031 (.76)	938 (.69)	4,555 (.67)

Note. [a] proportion of correct response. SF = small_fictitious; LF= large_fictitious; SS = small_small; LS = large_small; LL = large_large; F+_F+ = pairs in which both cities are more familiar; F_F = pairs in which both cities are less familiar; F+_F = pairs in which one city is more familiar and the other is less familiar; F+_U = pairs in which one city is more familiar and the other is unrecognized; F_U = pairs in which one city is less familiar and the other is unrecognized; U_U = pairs in which both cities are unrecognized

For the fictitious–fictitious (FF) pairs, there is no criterion to judge which one is the larger city, but participants may falsely recognize a city and choose accordingly. Among the total 1,356 FF pairs, 944 pairs resulted in the same level of familiarity. For the pairs of different familiarity ratings, participants chose the more familiar city more often (255 times out of 412, or 62%). A sign test revealed that participants were significantly more likely to choose the more familiar city, $z = 4.24, p < .00$, even among the fictitious items.

Model testing

As stated earlier, our adaptation of the *r-s* model in Fig. 2 uses the following pair classification: The familiarity-homogeneous tree (the first tree) includes: F_F (both objects recognized with low familiarity ratings), and F+_F+ pairs (both objects recognized with higher familiarity ratings). The familiarity-heterogeneous tree (the second tree) includes: F_F+ (both objects recognized but with different levels of familiarity ratings). The recognition tree (the third tree) includes: F_U (one recognized with a low familiarity rating and the other is unrecognized), and F+_U (one recognized with a high familiarity rating and the other is unrecognized). The guessing tree (the fourth tree) includes: U_U (both unrecognized). There were a total of 6,763 pairs, and of these, 1,048 were familiarity-heterogeneous cases. Frequency for each category can be found in Appendix Tables 7 and 8.

Model fitting was conducted using the MultiTree program (Moshagen, 2010), and the model in Fig. 2 was specified for each pair determined by population size SF (small–fictitious), LF (large–fictitious), LS (large–small), SS (small–small), LL (large–large) pair types in one step. Pairs FF were not included, as there is no criterion to judge which one is the larger city.

The model has 12 data categories (as shown in Fig. 2), and, given our design with five city-size pair types, there are 60 data categories, among which 43 are free. The number of free parameters is 40, resulting in three degrees of freedom for

tests. The model fitted the data well, $G^2(3) = 4.41, p = .22$. Estimated parameters appear in Table 3.

We note that in the present model *b3* is defined as in the original *r-s* model and equal to the weighted combination of parameters *b1* and *b2*, and the weight, *p*, is the proportion of cases judged as homogenous (both items have equal familiarity) relative to all cases with familiarity in which knowledge may be used (i.e., homogeneous and heterogeneous cases as defined here). When the weight *p* = 0, then *b3* = *b2*, implying that the knowledge validity is the same for all familiarity difference cases (i.e., the F_F+, F_U, and F+_U cases). At the other extreme, when

p = 1, then *b3* = *b1*, implying that the knowledge validity is the same for homogeneous and for cases in which only one object is recognized. Furthermore, if *b1* = *b2*, then *b3* is equal to either *b1* or *b2*, implying that the knowledge validity is the same irrespective of familiarity differences. The importance of presenting these equivalences is to show that the type of knowledge used when confronted with a highly familiar and an unrecognized object may be similar, or different from the knowledge used when both objects are recognized (with equal or different levels of familiarity). Because the knowledge recalled about a fictitious city is unlikely to be similar to the knowledge recalled about a real city, we freely estimated parameter *b3* for the SF and LF pairs.

Results in Table 3 demonstrate that the recognition validity (*a*) greatly varies across the pair types of real cities, with the LS type having the largest value, whereas it is at chance level for the SS type. This is an expected result because recognition is correlated with city size. More generally, the recognized city is more likely to be the larger city in the LF, SF, and LS pair types, whereas in the SS and LL pair types, the recognized city comes from the same city pool size as its partner. Consequently, the cue validity for the SS and LL pair types is much lower than that of the other pair types.

Similar to *a*, the knowledge validity (*b1*, *b2,* and *b3*) also varied across pair types, with values ranging from .48 to .84

Table 3   The *r-s* model fit statistics and parameter estimates of Experiment 1

Parameter	Psychological meaning	Estimate (SE)				
		SF	LF	SS	LS	LL
a	Recognition validity	.79 (.02)	.91 (.01)	.49 (.02)	.85 (.02)	.68 (.02)
b1	Knowledge validity, familiarity-homogeneous knowledge case	.58 (.04)	.67 (.04)	.52 (.03)	.71 (.02)	.73 (.02)
b2	Knowledge validity, familiarity-heterogeneous knowledge case	.65 (.05)	.84 (.04)	.48 (.06)	.82 (.04)	.75 (.05)
b3	Knowledge validity, given that only one object is recognized	.72 (.03)	.78 (.04)	.51[a]	.75[a]	.74[a]
c	Familiarity validity	.78 (.04)	.88 (.02)	.49 (.04)	.83 (.02)	.69 (.03)
g	Correct guessing	.54 (.02)	.53 (.03)	.55 (.02)	.59 (.03)	.51 (.04)
p	Proportion of familiarity-homogeneous knowledge cases	–	–	.55 (.02)	.51 (.02)	.64 (.02)
r	RH use (considering the recognition cue in isolation)	.42 (.05)	.56 (.05)	.60 (.04)	.60 (.05)	.65 (.04)
s	FH use (considering familiarity in isolation)	.27 (.11)	.41 (.10)	.57 (.06)	.53 (.07)	.57 (.05)

Note. [a] This number is computed and thus reported without a standard error. SF = small_fictitious; LF = large_fictitious; SS = small_small; LS = large_small; LL = large_large

(median equal to .69). People have more valid knowledge for real city pairs involving an L city (LS, and LL) than real city pairs involving an S city (SS). This is because familiarity is correlated with city size, hence people are more familiar with L cities and have more accurate knowledge than for S cities. The same is true for the SF and LF pair types. The parameters *b1*, *b2*, and *b3* are all larger for the LF than for the SF type. The results also show that *p* is not extreme, thus leads to values of *b3,* which are a balanced combination of the knowledge validity *b1* and *b2* for real city pair types. The estimated values of *b3* for the pairs with a fictitious city (.72 and .78) show that the knowledge used in those cases tended to be valid. Furthermore, having *p* not near 100% indicates that many cases with equal recognition have nevertheless unequal familiarity (e.g., heterogeneous cases such as F_F+ are both familiar, and thus both are recognized in traditional analysis of the RH but are heterogeneous in familiarity in the present analysis). This is important because we argue that a familiarity difference is key in making inferences even when both objects would be otherwise classified as recognized.

Both the pattern and value of parameter *c* (representing the cue validity of the familiarity difference) are very similar to that of *a* (the cue validity of the recognition heuristic). That is, *c* is larger for the SF, LF, and LS pair types and smaller for the SS and LL pair types. Focusing on the real city pairs, we compared parameters *a* and *c* by setting *a* = *c* in a reduced model. The result showed that there was no significant difference between *a* and *c* in these comparisons, $G^2(1) = .04$, *p* = .84; $G^2(1) = .54$, *p* = .46; $G^2(1) = .04$, *p* = .84, for the SS, LS, and LL types, respectively. In other words, the recognition cue validity for recognized–unrecognized pairs is the same as the familiarity difference cue validity for recognized–recognized cases with heterogeneous familiarity.

With regard to recognition, in the real pair types the *r* parameters are greater or equal to .60, which is congruent with previous findings in the literature. This result indicates that people rely on the RH cue in a majority of cases, even when the cue validity is as low as chance level. In addition, *r* of the SS pair type is the same as that of the LS pair type. Note that the cue validity (*a*) for the former is much lower than for that of the latter. In other words, higher cue validity is not associated with more frequent use of the recognition heuristic.

The parameter *s*, which represents the rate of using familiarity differences, was much higher than .25, which is the average rate of using the fluency heuristic as found in prior studies. Additionally, model comparisons revealed that there was no significant difference between *r* and *s*: $G^2(1) = .20$, *p* = .65; $G^2(1) = .85$, *p* = .36; $G^2(1) = 1.21$, *p* = .27, for the SS, LS, and LL types, respectively. This result indicates that the rate of using familiarity differences was similar to the rate of using the recognition cue.

Discussion

Experiment 1 demonstrated support for the application and generalization of the *r-s* model to inferential judgments with familiarity differences. The model fitted the data well and the parameters were interpretable. Self-reported familiarity was found to have high ecological validity as measured by correlation of familiarity ratings with city populations. Correct responding varied as a function of both familiarity and population size differences as defined by the experimental pair types.

Results of the experiment also replicated previous findings that the RH is used in a majority of cases. By fitting the model across different pair types in a single domain, we also found that the use of the RH was independent of recognition validity. That is, even when the cue validity was as low as chance level, people still used the RH in a substantial number of cases. In addition, when both cities were recognized but had a large

familiarity difference, people relied on that familiarity difference in more than 50% of the cases, as measured by parameter *s*. This rate is much larger than that of the fluency heuristic (Hilbig et al., 2011; Schwikert & Curran 2014), suggesting that it may be easier to rely on feelings of familiarity than on retrieval speed to make judgments.[3]

Results further showed that the proportion of cases that were familiar (and hence recognized) were not all of equal familiarity. This in turn yielded an almost even classification of homogeneous and heterogeneous cases, as measured by the parameter *p*. Furthermore, a subset of heterogeneous cases (with unequal familiarity) triggered the use of further knowledge, and this knowledge tended to be valid in most cases. The knowledge validity was also greater for comparisons including a large city than a small city, and this means that the knowledge people have of large cities is more accurate than that of small cities. We also note that the knowledge validity for cases including one fictitious city was high, and although individuals chose higher familiar cities more often than less familiar cities, even among fictitious–fictitious cases, the proportion of false recognition was low.

Another result is that the values of parameters *c* (the cue validity of the familiarity difference) and *a* (the cue validity of the recognition heuristic) were similar. They were both larger for the SF, LF, and LS pair types than for the SS and LL pair types. This means that using a heuristic for pairs with similar objective populations was less accurate.

Experiment 2 was designed to replicate these findings, in particular that the RH and the familiarity difference cue are used frequently. In addition, their status as fast heuristics is examined with the aid of response-time analysis. Specifically, assuming that people make faster decisions when using heuristics, the mean response time for heuristic use cases should be shorter than that of others.

Experiment 2

Experiment 2 differs from Experiment 1 in two aspects: the tasks were completed in a computer, and the response time was measured for the choice task. Congruent with previous findings that decisions are faster when they are in line with the RH, we expected to find that decisions made with heuristics (either the RH or the familiarity difference), as described by the *r-s* model, will be faster. More specifically:

- Hypothesis 1: *In the r-s model, the response times for cases classified in the familiarity-heterogeneous and the recognition trees (that is, cases for which the familiarity difference cue or the recognition cue is available to use, i.e. F+_F, F+_U, and F_U pairs) are shorter than that of cases classified in the familiarity-homogeneous tree (i.e. F+_F+, and F_F pairs).* That is, for cases in which a familiarity difference or RH cue is available to use, people can use them in isolation (in a majority of cases as estimated by the *r-s* model) and make quicker decisions.

- Hypothesis 2: *Furthermore, among cases classified into the familiarity-heterogeneous and the recognition trees in the r-s model (i.e., cases for which the familiarity difference cue or the recognition cue is available to use), the response times will be shorter for LL and LF pair types than for SS and SF pair types.* This hypothesis is in line with previous findings that congruent knowledge will lead to shorter response times. More specifically, having valid knowledge about the recognized (or the more familiar) city for the LL and LF pairs implies that the city is an L city, which is congruent with the implication of the recognition or the familiarity difference cue. In contrast, the SS and SF pair types can potentially bring ambiguity in responding because identifying a small city does not unequivocally guide choice. Consequently, the response time for the LL and LF pair types will be shorter than that of the SS and SF pair types.

Method

Materials

Experiment 2 used the same stimuli pairs as Experiment 1, presented via a computer program.

Participants

Fifty psychology undergraduate students at Ohio University participated in this experiment in exchange for a course credit.

Procedure

The experiment consisted of three parts, all presented via a computer program. Both the first and the third parts were familiarity ratings of cities in Ohio, randomly presented across participants. A 0–100 Likert-type rating scale was used, aiming to differentiate more subtle differences in familiarity ratings, with 0 = *no familiarity at all*, and 100 = *full familiarity*.

The second part was the city-size choice task. Computer programing was used so that each participant was presented randomly with 72 pairs, with 12 pairs of each of the pair types—FF, SF, LF, SS, LL, and LS. The two cities in each

---

[3] One difference is that the FH classified more cases into the heterogeneous knowledge pairs (lower *p* in the *r-s* model) than did the familiarity difference cue in this study. It is possible that many of the classified FH differences were too low to be used in choice by the participants in the studies of Hilbig et al. (2011) and Schwikert and Curran (2014).

pair were selected at random from their respective types. Using this procedure, there were 576 (24 × 24) ways to construct the pair types, and each of them had the same probability of being presented to the participants. In order to ensure that all cities were presented to each participant, each of the 72 cities was used two times to construct the 72 pairs for each participant. In addition, the pairs were presented in such a way that for every 12 presentations, there were two presentations of each of the six pair types.[4]

Before the presentation of a city pair, a "+" was shown in the center of the computer screen. After 150 milliseconds, the "+" disappeared, and two city names appeared in the middle of the screen. As quickly as they could, participants had to press "Z" if they thought that the left city had the larger population and press "M" if they thought that the right city had the larger population. Once the participant pressed one of the two keys, the symbol "+" displayed again. The next trial then followed. The data for this experiment included both the response time and the choices the participants made.

Results

Classification of city types

As was true in Experiment 1, there was no significant difference on the mean of familiarity ratings between the first and second tests, so we used the ratings from the first task in the following analysis. For all ratings across cities and participants, 51.4% of the cases received a rating of 0, or no familiarity, and these were classified as unrecognized (U). Among the rest, there were 24.4% familiarity ratings that fell below 50, while the remaining percentages were above 50. The less familiar type (F type) was created with ratings greater than 0 but below 50, and the rest were classified as the more familiar type (F+ type).

Correlation between city size and familiarity rating

Similar to Experiment 1, there was a significant correlation between city size and familiarity rating computed across participants, $r(46) = .84$, $p < .001$. At the individual level, 49 out of 50 (98%) participants showed a positive correlation between city sizes and familiarity ratings.

Correct responding

Performance varied with both familiarity and population size types as shown by a significant chi-square test, $\chi^2(20) =$

491.21, $p < .00$. Comparisons including a fictitious city usually lead to correct answers (71% and 79% correct for SF and LF, respectively). As in Experiment 1, the rates of selection of fictitious cities were not large, but neither were they negligible (i.e., 29% and 21%, respectively). Across familiarity judgment pair types, participants performed worst for the SS type. Across population size types, participants performed best for the F+_U type followed by the F_U and F+_F types (see Table 4).

Note that among the total 610 FF pairs (not shown in Table 4), a great majority of cases (416 pairs, equivalent to 68.2%) had the same familiarity. For the pairs with different familiarity ratings, the tendency was to select the more familiar city in 104 of the 194 cases (53%). However, a sign test revealed that there was not a significant preference in selection, $z = 1.10$, $p = .27$.

Model fitting

Similar to Experiment 1, the familiarity-homogenous tree in the r-s model includes the F_F and F+_F+ types; the familiarity-heterogeneous tree includes the F_F+ type; the recognition tree includes the F_U and F+_U types; and the U_U type was classified into the fourth node.

We classified the total 2,990 pairs according to the r-s model. There were 420 familiarity-heterogeneous cases (see Appendix Tables 7 and 8 for more detailed frequency). The number of cases in some data categories (for the SF and LF pair types) were less than five, and thus we had to exclude these pairs from the analyses. As done in Experiment 1, the FF type was also excluded as there is no correct choice for this type. The r-s model fitted the data well, $G^2(3) = 1.23$, $p = .75$, and estimated parameters appear in Table 5.

Results show that the recognition validity (a) is lowest for the SS type and largest for the LS type, replicating the finding of Experiment 1. That is, cue validity was smaller when the two comparative cities came from the same city pool than when they came from two different city pools.

The familiarity difference cue validity (c) and the recognition cue validity (a) demonstrated similar patterns. The validity, c was larger for the LS than for the SS. Setting $a = c$ resulted in a nested model that was not significantly different from the full model in each of the SS, LS, and LL pair types, $G^2(1) = 2$, $p = .16$; $G^2(1) = 2.76$, $p = .10$; $G^2(1) = .59$, $p = .44$, for the SS, LS, and LL pair types, respectively. This result indicates that the heuristics have comparable validity.

The percentage of familiarity homogeneous cases among all knowledge cases (p) ranged from .59 to .71 in Experiment 2. This is slightly higher than that of Experiment 1, but still yielded a good proportion of cases in which choice may be made on the basis of higher familiarity. In both experiments p is highest for the LL type.

---

[4] After the experiment, we found that each participant was presented with 72 pairs, but the number of each pair type varied from 10 to 14 instead of a stable 12. This may have been caused by a programing error. In total, there were 3600 pairs with FF (610), FS (602), FL (593), SS (596), SL (600), and LL (599).

Table 4 Frequencies of correct response of familiarity judgment type by population size type

	SF	LF	SS	LS	LL	Total
F+_F+	4 (.67[a])	8 (.80)	26 (.49)	63 (.65)	149 (.75)	250 (.68)
F_F	41 (.72)	35 (.63)	24 (.39)	29 (.62)	35 (.66)	164 (.59)
F+_F	42 (.84)	68 (.84)	39 (.57)	76 (.77)	66 (.64)	291 (.72)
F+_U	65 (.81)	170 (.93)	55 (.50)	122 (.81)	71 (.72)	483 (.78)
F_U	98 (.83)	100 (.89)	46 (.44)	76 (.77)	51 (.78)	371 (.75)
U_U	176 (.60)	85 (.56)	111 (.55)	63 (.58)	50 (.63)	485 (.58)
Total	426 (.71)	466 (.79)	301 (.51)	429 (.72)	422 (.70)	2,044 (.68)

Note. [a] proportion of correct response. SF = small_fictitious; LF = large_fictitious; SS = small_small; LS = large_small; LL = large_large; F+_F+ = pairs in which both cities are more familiar; F_F = pairs in which both cities are less familiar; F+_F = pairs in which one city is more familiar and the other is less familiar; F+_U = pairs in which one city is more familiar and the other is unrecognized; F_U = pairs in which one city is less familiar and the other is unrecognized; U_U = pairs in which both cities are unrecognized

Similar to a, the knowledge validity ($b1$, $b2$, and $b3$) varied across pair types (they ranged from .44 to .8), with larger values for pairs involving an L city. That is, people were more familiar with L cities and have more accurate knowledge than for S cities.

As in Experiment 1, people used the recognition heuristic in a majority of cases as revealed by the values of parameter r. Similar to Experiment 1, although the cue validity (a) was highest for the LS pair type, the rate of using the RH (r) was not greater for this type than for that of the LL pair type. This finding indicates that a larger cue validity was not associated with a more frequent use of the recognition heuristic.

Overall, the rate of using the familiarity difference cue (s) ranged from .43 (LS pair type) to .65 (SS pair type), which was greater than the average 25% found for the fluency heuristic in prior studies. Comparing the values of r and s, we found that there were no significant differences for the SS and LS pair types, $G^2(1) = 0.6$, $p = .44$; $G^2(1) = 2.57$, $p = .11$, respectively. For the LL type, the value of r was

significantly higher than s, $G^2(1) = 5.08$, $p = .02$. This finding indicates that the use of the RH was higher for pairs of large cities.

Response-time analysis

To test the first hypothesis, response time for each trial was log transformed. For each participant, a median response time was obtained for each of the F+_F+, F_F, F_F+, F_U, F+_U, and U_U categories across population pair types. The means of response times across participants are shown in Table 6. A repeated-measures ANOVA revealed that there was a significant difference in response time as a function of data categories, $F(5, 16) = 10.16$, $p < .001$. A contrast test revealed that the average response time of familiarity-homogeneous cases (F+_F+, F_F) was significantly longer than that of heterogeneous and recognition cases (F+_U, F_U, and F+_F), $F(1, 20) = 5.445$, $p < .03$, hence supporting Hypothesis 1.

To test Hypothesis 2, for the 2,113 cases classified into the familiarity-heterogeneous and recognition trees, the median

Table 5 The r-s model fit statistics and parameter estimates of Experiment 2

		Estimates (*SE*)		
Parameter	Psychological meaning	SS	LS	LL
a	Recognition validity	.47 (.04)	.82 (.02)	.74 (.03)
$b1$	Knowledge validity, familiarity-homogeneous knowledge case	.44 (.04)	.65 (.04)	.73 (.03)
$b2$	Knowledge validity, familiarity-heterogeneous knowledge case	.56 (.11)	.80 (.06)	.58 (.10)
$b3$	Knowledge validity, given that only one object is recognized	.49[a]	.71[a]	.69[a]
c	Familiarity validity	.57 (.06)	.75 (.04)	.70 (.05)
g	Correct guessing	.56 (.03)	.58 (.05)	.63 (.05)
p	Proportion of familiarity-homogeneous knowledge cases	.62 (.04)	.59 (.03)	.71 (.02)
r	RH use (considering the recognition cue in isolation)	.56 (.06)	.62 (.06)	.80 (.05)
s	FH use (considering familiarity in isolation)	.65 (.09)	.43 (.10)	.59 (.09)

Note. [a] This number is computed and thus reported without a standard error. SS = small_small; LS = large_small; LL = large_large

Table 6 Mean response time in milliseconds

Choice type	Mean	N
F+_F+	1,301.51	37
F_F	1,277.04	37
F+_F	1,160.65	43
F+_U	1,120.80	44
F_U	1,223.92	45
U_U	1,363.91	44

Note. N is the number of participants with choices belonging to the respective category. F+_F+ = pairs in which both cities are more familiar; F_F = pairs in which both cities are less familiar; F+_F = pairs in which one city is more familiar and the other is less familiar; F+_U = pairs in which one city is more familiar and the other is unrecognized; F_U = pairs in which one city is less familiar and the other is unrecognized; U_U = pairs in which both cities are unrecognized

values of the log transformed response times were obtained for the SS, LL, SF, LF, and LS pair types. A repeated-measures ANOVA on this variable revealed that there was a significant difference among population pair types, $F(4, 38) = 13.41, p < .001$. A contrast test revealed that the average of the median response times for the SS and SF pair types was significantly longer than that of the LF and LL pair types, $F(1, 41) = 33.21, p < .001$, supporting Hypothesis 2. That is, when the more familiar city was S it takes longer to choose it than when the more familiar city was L.

Discussion

Experiment 2 replicated and expanded the results of Experiment 1. The ecological validity of the familiarity difference cue was high, correct responding varied with both familiarity and city size differences, and the *r-s* model fitted the data well. For the RH cue, the proportion of choosing the recognized city was similar to that found in previous studies (Hilbig et al., 2010; Hilbig et al., 2011; Schwikert & Curran, 2014). When both cities were recognized but with a large familiarity difference, the ratio of using the familiarity difference cue ranged from 45% to 65%, again much higher than the average of 25% rate of using the fluency heuristic in previous studies (Hilbig et al., 2011; Schwikert & Curran, 2014).

Analyses of the model parameters replicated the findings of Experiment 1 by first demonstrating that a substantial number of recognized cases could be further classified as heterogeneous based on familiarity differences as shown by values of parameter *p*. Second, the validity of the recognition and the familiarity difference cues varied with city-size pair types, and were larger for pair types including an L city. The knowledge validity parameters were also larger for pair types including an L type.

Experiment 2 measured response times, and this allowed us to test hypotheses regarding the processing time. Congruent with the categorization of objects and processing states of the *r-s* model, the response time was significantly shorter for pairs from the familiarity-heterogeneous and the recognition trees than for pairs from the familiarity-homogeneous tree. That is, as estimated by the *r-s* model, for pairs from the heterogeneous and recognition trees, people used heuristics by ignoring other knowledge in the majority of cases, which sped the decision-making process. In contrast, for pairs from the homogenous tree, people could not rely on a simple heuristic and had to make selections based on whatever knowledge they could bring to the task; this in turn slowed down the decisions they made.

Furthermore, within the heterogeneous cases and recognition cases, we found that the processing time was significantly shorter in the pairs for which the more familiar city was large rather than small. These results suggest that even when further knowledge is called upon, response time is significantly shorter when the knowledge is congruent with the cues than when they are in conflict with each other.

General discussion

The main thesis of this project is that familiarity differences play a key role in inferential judgments in the city-size task. In the spirit of Schwikert and Curran (2014), we derived a variant of the *r-s* model based on self-perceptions of city familiarity and explored several aspects of the inferential judgment process. The current study showed that the familiarity difference cue was used in a large proportion of cases (more than 50% of the time) and that selecting by familiarity was generally a valid strategy. The response-time analysis supported the claim that the use of the familiarity difference was fast. In other words, when two comparative items were both recognized, a familiarity difference could be and was often used in a fast-and-frugal manner.

Key results from our experiments were an indication that correct responding varied with pairs of cities created with population differences, as well as with pairs of cities determined by judgments of familiarity. With the aid of the *r-s* model, the pattern of responding could be explained by parameters representing different aspects of the judgment and decision processes. Both experiments showed high usage and high validity for the studied heuristics. The experiments also demonstrated higher knowledge validity for pair types containing a large city. This implies that when individuals use knowledge to make a selection, knowledge tends to be better when a large city is in the comparison. That is, individuals tend to be more familiar with cities with larger population and also have greater valid knowledge for them.

The proposition and findings that familiarity differences are an intricate aspect of inferential judgments is consistent with Castela et al.'s (2014) construct of memory strength, and

with their results based on the classification of highly recognized and merely recognized objects. As stated, our results also showed that the rate of use of the familiarity difference cue was greater than the rate of use of the fluency heuristic in previous studies. This lower rate of fluency heuristic usage cannot be attributed to its availability. Indeed, the fluency heuristic was available in a higher proportion of cases in Hilbig et al. (2011) and in Schwikert and Curran's (2014) studies (77% and 70%, respectively) compared to the proportion of heterogeneous cases in the current experiments (an average of 40%). However, the study conducted by Hilbig et al. (2011) showed a lower fluency heuristic usage even with a higher fluency threshold. They found a maximum rate of 31% among 11 different fluency thresholds that ranged from 0 to 1,000 milliseconds.

We attribute the observed discrepancy in the heuristic rate's use when comparing familiarity and fluency to the relative ease of using self-perceptions of familiarity. Compared to fluency, self-evaluation is easier to observe than fluency, which is covert. The study by Schwikert and Curran (2014) sheds some light on this discrepancy. These authors measured event-related potentials while participants were conducting city-size tasks. They found that there was no significant difference in FN400 (an event-related potential assumed to index familiarity) between faster and slower recognized cities. In other words, assuming the FN400 is measuring familiarity in some awareness kind of level (but see Leynes, Bruett, Krizan, & Veloso, 2017, for a different view), they found no difference in familiarity when fluency differences existed. This may be interpreted to indicate that fluency is not related to memory strength. Thus, explicit familiarity ratings are probably more likely to be related to actual memory strength, but future studies using FN400 are needed to verity this possibility. Additionally, covert fluency differences may just be harder to detect and guide choices than overt, self-reported feelings of familiarity. In some sense, feelings of familiarity may be more available as information that can influence people's decisions in a more direct manner.

Schwikert and Curran (2014), however, interpreted their findings in a different manner. They proposed that the fluency heuristic may guide choice when the difference between subjective feelings of familiarity, as well as differences in recollection knowledge, are not present. In other words, the fluency cue may be used only when both familiarity differences and knowledge are unavailable to aid decisions. Unfortunately, the previous studies on the fluency heuristic (Hilbig et al., 2011; Schwikert & Curran 2014) assessed its rate of usage among knowledge cases that included pairs with small familiarity differences but also pairs with relatively large familiarity differences. Thus, the rate of using fluency when familiarity differences are absent is not known.

An additional consideration when contrasting fluency with self-perceptions of familiarity is the finding that there is an inverse correlation between recognition time and self-reported subjective familiarity (Schweickart & Brown, 2014). More specifically, these researchers found that the more familiar an object is, the quicker the retrieval process. However, the inverse relation is statistical and imperfect; therefore, selections based on fluency differences may not always coincide with selections based on subjective feelings of familiarity differences. We thus fall back on Schwikert and Curran's (2014) findings on similar levels of brain activation indexing familiarity for differences in recognition speed, which indicates that fluency differences do not necessary indicate memory-strength differences. Nevertheless, studies are need explore the relationship of fluency and self-reported familiarity differences more precisely. More generally, future studies are needed to verify whether fluency aids judgment in cases where the familiarity and recollection knowledge differences are unavailable (as described in Fig. 1), or whether the fluency and familiarity difference heuristics coincide to influence choice.

The response time analysis in Experiment 2 provided supplemental evidence that the use of the familiarity difference cue, or the RH is fast. Note, however, fast response time does not necessarily imply a frugal decision-making process. As illustrated by Glöckner and Betsch (2012), additional knowledge could speed up a decision process, as long as the knowledge is coherent with other information. This is consistent with what was found in the current work, namely, that decisions can be faster when knowledge is congruent with a heuristic as compared to when they are in conflict. We emphasized that it is the response-time analysis, in combination with the r-s model analysis, that lead to the implication that a familiarity difference cue may be fast and frugal in some conditions.

Another finding worth noting is that, within the same domain, the ratio of using the RH did not vary with cue validity. This result is congruent with findings from the study conducted by Pohl (2006). In that study, Pohl found that the use of the RH was independent from cue validity at the individual level, showing that there was no correlation between individual cue validity and the percentage of using the RH. These results invalidate the proposal that people can adaptively adjust the probability of using a heuristic strategy according to cue validity within the same domain, as suggested by some researchers (e.g., Pachur & Hertwig, 2006). However, we note that this result cannot be generalized to tasks across domains. The study conducted by Horn, Pachur, and Mata (2015) found that the rate of using the RH was much higher in the standard city-size task (with a higher cue validity), than in an infectious disease task that asked participants to select the disease with higher incidence rate (which had a lower cue validity).

Regarding the effect of false recognition, Experiment 1 found that participants were more likely to choose the more familiar city in fictitious–fictitious pairs. Experiment 2 found the same trend, but it did not reach significance. To summarize, the present results agree with the assertion that the

familiarity difference cue, as a broader concept than the recognition cue, plays a fundamental role in making inferences (Schwikert & Curran, 2014). However, the framework by Schwikert and Curran (2014) assumes that a large familiarity difference cue is *always* used prior to using knowledge. Our results in the response time analysis showed that the response time was shorter when differences in familiarity (recognition) were available and knowledge was congruent with the heuristic. This implies that people consider knowledge at different levels of familiarity. The order of processing and the manner by which individuals incorporate knowledge is not known. Indeed, it is possible that a familiarity judgment brings a large amount of information to mind quickly, and that pairs with large familiarity differences are easier to judge because the more familiar object's informational base is also larger. We note that these findings are based on multinomial processing trees, which are good approximations of the cognitive processes and can provide meaningful assessments of strategy use, as shown here, but they rely on categorization of otherwise continuous internal states. Furthermore, the independence of processing stages may not always be justified.

The adjustable spanner theory proposed by Lee and Cummins (2004) provides an alternative processing account of the inferential judgment process. According to this theory, judgment and decision making is a process of accumulation of evidence in favor of one of the objects based on a sequential sampling of information. The accumulation of evidence continues until a threshold is reached. The recognition cue is regarded as one type of evidence, but this is similar to other knowledge. In other words, evidence comes in different

forms. In addition, there are individual differences regarding the level of the threshold (Newell, 2005). That is, even when other information is available for choice, people may stop looking for information when a threshold is already attained or exceeded. Accordingly, the theory also predicts that the response time will be shorter when the familiarity difference cue is in convergence with other knowledge. When the sampled information points in the same direction as the familiarity difference cue, then shorter response times ensue.

More research is needed to further explore accumulation type models in contrast to multinomial processing trees in order to advance a greater understanding of the knowledge/familiarity difference-cue incorporation in comparative judgments. In particular, we followed researchers who treated the familiarity difference cue as a categorical variable (Schwikert & Curran, 2014) rather than as a continuous variable (Honda et al., 2011). Future studies with alternative modeling approaches are needed to more fully address the role of familiarity differences in choice using the full spectrum of variability embedded in the continuous ratings.

Limitations

We note that participants made only 72 inferences across six pair types. This design resulted in having only few data points in some categories and therefore made the fitting of the *r-s* model at the individual level impossible. Future studies can test the model at the individual level to address individual difference in the use of heuristics.

Appendix

Table 7 The *r-s* model categories and observed choice frequency of Experiment 1

r-s model category	*r-s* model category meaning	SF	LF	SS	LS	LL
1	Familiarity-homogeneous knowledge case, correct judgment	82	67	116	222	383
2	Familiarity-homogeneous knowledge case, false judgment	59	33	106	89	137
3	Neither recognized, correct judgment	324	170	238	117	69
4	Neither recognized, false judgment	276	151	194	82	67
5	One recognized, adherence to RH, correct judgment	347	607	202	423	243
6	One recognized, adherence to RH, false judgment	64	46	212	57	100
7	One recognized, nonadherence to RH, false judgment	66	66	49	43	33
8	One recognized, nonadherence to RH, correct judgment	46	24	53	25	27
9	Familiarity-heterogeneous knowledge cases, adherence to familiarity difference cue, correct judgment	50	150	71	225	185
10	Familiarity-heterogeneous knowledge cases, adherence to familiarity difference cue, false judgment	10	11	74	31	60
11	Familiarity-heterogeneous knowledge cases, nonadherence to familiarity difference cue, false judgment	17	15	20	21	19
12	Familiarity-heterogeneous knowledge cases, nonadherence to familiarity difference cue, correct judgment	9	11	19	19	31

Table 8 The *r-s* model categories and observed choice frequency of Experiment 2

r-s model category	*r-s* model category meaning	SS	LS	LL
1	Familiarity-homogeneous knowledge case, correct judgment	50	92	184
2	Familiarity-homogeneous knowledge case, false judgment	64	52	69
3	Neither recognized, correct judgment	111	63	50
4	Neither recognized, false judgment	88	45	30
5	One recognized, adherence to RH, correct judgment	77	185	115
6	One recognized, adherence to RH, false judgment	90	30	35
7	One recognized, nonadherence to RH, false judgment	23	21	6
8	One recognized, nonadherence to RH, correct judgment	24	13	7
9	Familiarity-heterogeneous knowledge cases, adherence to familiarity difference cue, correct judgment	33	65	59
10	Familiarity-heterogeneous knowledge cases, adherence to familiarity difference cue, false judgment	24	14	24
11	Familiarity-heterogeneous knowledge cases, nonadherence to familiarity difference cue, false judgment	6	9	13
12	Familiarity-heterogeneous knowledge cases, nonadherence to familiarity difference cue, correct judgment	6	11	7

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Castela, M., Kellen, D., Erdfelder, E., & Hilbig, B. E. (2014). The impact of subjective recognition experiences on recognition heuristic use: A multinomial processing tree approach. *Psychonomic Bulletin & Review, 21*(5), 1131–1138. https://doi.org/10.3758/s13423-014-0587-4

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.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(4), 552–564.

Gigerenzer, G., & Goldstein, D. G. (2011a). The beauty of simple models: Themes in recognition heuristic research. *Judgment and Decision Making, 6*, 392–395.

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

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98*(4), 506–528.

Glöckner, A., & Betsch, T. (2012). Decisions beyond boundaries: When more information is processed faster than less. *Acta Psychologica*, *139*(3), 532–542.

Goldstein, D. G., & Gigerenzer, G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103*(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

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

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review, 109*(1), 75–90. https://doi.org/10.1037/0033-295x.109.1.75

Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning,*

Memory, and Cognition, *34*, 1191–1206. https://doi.org/10.1037/a0013025

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*(1), 123–134. https://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*(4), 827–839. https://doi.org/10.1037/a0022638

Hilbig, B. E., & Pohl, R. F. (2008). Recognizing users of the recognition heuristic. *Experimental Psychology*, *55*, 394–401. https://doi.org/10.1027/1618-3169.55.6.394

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*(5), 1296–1305. https://doi.org/10.1037/a0016565

Hilbig, B. E., Pohl, R. F., & Broder, A. (2009). Criterion knowledge: A moderator of using the recognition heuristic? *Journal of Behavioral Decision Making, 22*(5), 510–522. https://doi.org/10.1002/bdm.644

Honda, H., Abe, K., Matsuka, T., & Yamagishi, K. (2011). The role of familiarity in binary choice inferences. *Memory and Cognition*, *39*(5), 851–863. https://doi.org/10.3758/s13421-010-0057-9

Horn, S., Pachur, T., & Mata, R. (2015). How does aging affect recognition-based inference? A hierarchical Bayesian modeling approach. *Acta Psychologica*, *154*, 77–85. https://doi.org/10.1016/j.actpsy.2014.11.001

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology. 77*(6), 1121–1134.

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. https://doi.org/10.3758/BF03196581

Leynes, P. A., Bruett, H., Krizan, J., & Veloso, A. (2017). What psychology process is reflected in the FN400 event-related potential component? *Brain and Cognition, 113,* 142–145. https://doi.org/10.1016/j.bandc.2017.02.004

Moshagen, M. (2010). MultiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods, 42*, 42–54. https://doi.org/10.3758/BRM.42.1.42

Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Sciences, 9*(1), 11–15. https://doi.org/10.1016/j.tics.2004.11.005

Newell, B. R., & Fernandez, D. (2006). On the binary quality of recognition and the inconsequentiality of further knowledge: Two critical tests of the recognition heuristic. *Journal of Behavioral Decision Making, 19*(4), 333–346. https://doi.org/10.1002/bdm.531

Newell, B. R., & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning Memory and Cognition, 30*(4), 923–935. https://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*(1), B1–B9. https://doi.org/10.1016/s0010-0277(03)00141-0

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Science, 12*, 237–241. https://doi.org/10.1016/j.tics.2008.02.014

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 & Cognition, 32*(5), 983–1002. https://doi.org/10.1037/0278-7393.32.5.983

Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. G. (2011). The recognition heuristic: A review of theory and tests. *Frontiers in Psychology, 2*, 1–14. https://doi.org/10.3389/fpsyg.2011.00147

Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal-detection models: Comment on Wixted. *Psychological Review, 114*(1), 188–201. https://doi.org/10.1037/0033-295x.114.1.188

Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making, 19*(3), 251–271. https://doi.org/10.1002/bdm.522

Pohl, R. F., Erdfelder, E., Michalkiewicz, M., Castela, M., & Hilbig, B. E. (2016). The limited use of the fluency heuristic: Converging evidence across different procedures. *Memory & Cognition, 44*(7), 1114–1126. https://doi.org/10.3758/s13421-016-0622-y

Reid, A., & González-Vallejo, C. (2009). Emotion as a tradeable quantity. *Journal of Behavioral Decision Making, 22*(1), 62–90. https://doi.org/10.1002/bdm.612

Richter, T., & Spath, P. (2006). Recognition is used as one cue among others in judgment and decision making. *Journal of Experimental Psychology: Learning Memory and Cognition, 32*(1), 150–162. https://doi.org/10.1037/0178-7393.32.1.150

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

Schweickart, O., & Brown, N. R. (2014). Magnitude comparison extended: How lack of knowledge informs comparative judgments under uncertainty. *Journal of Experimental Psychology: General, 143*, 273–294. https://doi.org/10.1037/a0031451

Schwikert, S. R., & Curran, T. (2014). Familiarity and recollection in heuristic decision making. *Journal of Experimental Psychology: General, 143*(6), 2341–2365. https://doi.org/10.1037/xge0000024

Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 1235–1253.

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29*, 716–732.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning Memory and Cognition, 20*(6), 1341–1354. https://doi.org/10.1037/0278-7393.20.6.1341

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology. 9*(2), 1–27. https://doi.org/10.1037/h0025848