Memory-based Decision Making: Familiarity and Recollection in the Recognition and Fluency Heuristics

Shane Ross Schwikert
University of Colorado at Boulder, shane.schwikert@colorado.edu

Follow this and additional works at: https://scholar.colorado.edu/psyc_gradetds

Part of the Cognitive Psychology Commons, and the Neurosciences Commons

Recommended Citation
https://scholar.colorado.edu/psyc_gradetds/83

This Thesis is brought to you for free and open access by Psychology and Neuroscience at CU Scholar. It has been accepted for inclusion in Psychology and Neuroscience Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
Memory-based Decision Making: Familiarity and Recollection in the Recognition and Fluency Heuristics

Shane R. Schwikert
B.S., University of Michigan, 2009
Advisor: Tim Curran

A thesis submitted in partial fulfillment of the requirements for the degree Master of Arts in Cognitive Psychology at University of Colorado Boulder
September, 2013

Author Note
Shane R. Schwikert, Department of Psychology and Neuroscience, University of Colorado Boulder.

Correspondence concerning this article should be addressed to Shane R. Schwikert, Department of Psychology and Neuroscience, University of Colorado Boulder, Muenzinger Psychology, D434, Boulder, Colorado 80309-0345. E-mail: shane.schwikert@colorado.edu
This thesis entitled:
Familiarity and Recollection in Memory-based Decision Making
by Shane R. Schwikert
has been approved for the Department of Psychology and Neuroscience

___________________________________
Tim Curran

___________________________________
Alice Healy

___________________________________
Randy O’Reilly

Date: 10/2/2013

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline

HRC protocol # 11-0504
FAMILIARITY AND RECOLLECTION IN DECISION MAKING

Schwikert, Shane R. (M.A., Psychology and Neuroscience)

Memory-based Decision Making: Familiarity and Recollection in the Recognition and Fluency Heuristics

Thesis directed by Professor Timothy Curran

Abstract

Simple heuristics have been shown to facilitate the interplay between memory and judgment processes by exploiting fundamental cognitive abilities. The recognition and fluency heuristics are prime examples of shortcuts that capitalize on the by-products of memory retrieval to make quick decisions. In Experiment 1, we used a city-size comparison task while recording event-related potentials (ERPs) to investigate the potential contributions of familiarity and recollection to the two heuristics. ERPs were markedly different for recognition heuristic-based decisions and fluency heuristic-based decisions, suggesting a role for familiarity in the recognition heuristic and recollection in the fluency heuristic. In Experiment 2, we coupled a city-size comparison task with measures of perceived pre-experimental memory for each stimulus in the task. We found that more speedily recognized regions were also associated with greater amounts of recollection. Although previous literature suggests the fluency heuristic relies on recognition speed alone, our results suggest differential contributions of recognition speed and recollected knowledge to fluency heuristic-based decisions, whereas the recognition heuristic relies on familiarity.
# CONTENTS

## I. INTRODUCTION

The Recognition Heuristic ........................................3
The Fluency Heuristic ...........................................8

## II. EXPERIMENT 1

Methods ........................................................................16
  Participants ..............................................................16
  Materials and Procedure ........................................16
  EEG/ERP Methods ................................................19
Results .........................................................................20
  Behavioral Analyses ..............................................20
  ERP Analyses ........................................................26
Discussion ....................................................................33

## III. EXPERIMENT 2

Methods ..........................................................................41
  Participants ..............................................................42
  Materials and Procedure ........................................42
Results ..........................................................................44

## IV. GENERAL DISCUSSION

The Recognition Heuristic ........................................56
The Fluency Heuristic ................................................58
Theoretical Implications for the RH and FH ..................62
Conclusions ...............................................................69

REFERENCES ..................................................................71

APPENDIX

A. Experiment 1: r-s-Model Categories and Observed Choice Frequencies ........................................76

B. Experiment 2: r-s-Model Categories and Observed Choice Frequencies ........................................76
TABLES

Table

1. Experiment1: r-s-Model Output ................................................................. 24
2. Experiment2: r-s-Model Output ................................................................. 45
FIGURES

Figure

1. The r-s-Model .................................................................12
2. Sample trial from the population inference task.........................19
3. HydroCel Geodesic Sensor Net™ ......................................27
4. Experiment 1: Event-Related Potentials (ERPs)..........................28
5. Experiment 1: Topographic maps of amplitude differences......... 28
6. Experiment 1: Bar graph of mean ERP amplitude differences........ 29
7. Experiment 2: Individual probabilities of RH-use: Adherence rates vs. r-parameters.................................................................47
8. Experiment 2: Individual probabilities of FH-use: Adherence rates vs. s-parameters.................................................................48
9. Experiment 2: Boxplot of aggregate reaction times (RTs) for different levels of perceived memory.................................................49
10. Experiment 2: Inference choices for fluency heuristic (FH) trials based on perceived memory differences...........................................50
Memory-based Decision Making: Familiarity and Recollection in the Recognition and Fluency Heuristics

The study of how people make judgments and come to decisions has often acknowledged a role of memory in shaping these processes. For example, the fast-and-frugal heuristics research program (e.g. Gigerenzer, 2004) promotes an adaptive toolbox approach, suggesting that the mind has any number of specific heuristic judgment rules it can apply in conditionally appropriate situations. Many of these heuristics, notably the recognition heuristic and the fluency heuristic, are presumed to rely directly upon memory processes to make a judgment. The recognition heuristic is said to rely simply on recognition of objects to make quick choices, whereas the fluency heuristic is said to rely on recognition speed, or the speed of retrieval from memory to make choices. However, there has been an underappreciation in the heuristics research program for the specific underlying memory processes that presumably enable these heuristics to function. Alternatively, there has been little work done from a memory perspective to extend current theories of memory to heuristic decision-making. The current study aims to map the widely accepted dual-process account of recognition memory (Diana & Reder, 2006; Rugg & Curran, 2007; Wixted, 2007; Yonelinas, 2002) onto the recognition and fluency heuristics.

In an early article that could be considered the basis for the fast-and-frugal heuristics, research program, Gigerenzer, Hoffrage, and Kleinbölting (1991) discussed the potential role of recognition in particular when placing bets about unknown properties of the environment. Since then, two distinct but related heuristics have evolved that are presumed to rely heavily, if not entirely, on recognition memory. The first, coined the recognition heuristic (RH) by Goldstein and Gigerenzer (2002), was proposed for two-alternative forced choice tasks where one has to decide which of two objects scores higher on a given criterion. Consider the task of judging
which of two objects has the higher value, for example, which city is more populous: Boston or Saugatuck? The RH posits that if exactly one of these two items is recognized, then this item should be inferred to have the higher criterion value. So if one recognizes Boston but does not recognize Saugatuck, the RH predicts that Boston should be chosen as being more populous.

Inherent in the RH’s definition is its conditional use - it can only be applied when one item is recognized and one item is not recognized. Consequently, when both items are recognized the decision maker must resort to an alternate strategy (if we adopt the adaptive toolbox approach promoted by the fast-and-frugal heuristics program). Formalized by Schooler and Hertwig (2005), the fluency heuristic (FH) posits that if both items within a pair are recognized, one should compare the recognition speeds, or retrieval times, of both items and infer that the item retrieved more quickly from memory has the higher criterion value. Returning to the city-size comparison task, if one recognizes both Boston and Tulsa, but they retrieve Boston more quickly from memory, then the FH posits that Boston should be chosen as being more populous. The FH builds on a vast and diverse body of work on fluency (Begg, Anas, Farinacci, 1992; Hertwig, Gigerenzer, & Hoffrage, 1997; Jacoby & Dallas, 1981; Jacoby, Kelley, Brown, & Jasechko, 1989), much of which conceptualizes fluency in different ways. The particular theory we are concerned with is Schooler and Hertwig’s (2005) FH theory, where retrieval fluency is defined as how long it takes to retrieve a trace from long-term memory, or the speed with which objects are judged to be recognized.

Although all accounts of the RH assume it relies on a binary recognition judgment, few studies have probed the underlying memory processes that determine this recognition judgment. Likewise, the fast-and-frugal heuristics literature has largely ignored how underlying memory processes may influence retrieval fluency in the FH. Results from behavioral/cognitive,
neuropsychological, and neuroimaging studies of human memory increasingly indicate that recognition memory performance reflects two distinct memory processes or types of memory, often referred to as *familiarity* and *recollection* (Rugg & Curran, 2007; Rugg & Yonelinas, 2003; Woodruff, Hayama, & Rugg, 2006; Yonelinas, 2002). Familiarity-based recognition is considered fast-acting, relatively automatic, and does not involve the retrieval of qualitative information about an encoding episode. By contrast, recollection is conceived as a slower, more effortful process that gives rise to consciously accessible contextual information about an item. We will review the potential contributions of familiarity and recollection to each heuristic, starting with the recognition heuristic.

**The Recognition Heuristic**

There have been few attempts in the literature to directly parse out the contributions of familiarity and/or recollection to RH-based decisions. However, several theoretical claims surrounding the RH point to familiarity as being the primary mechanism driving the heuristic. In an article pre-dating their formalization of the RH, Gigerenzer et al. (1991) claimed that in a task in which one has to decide which of two objects scores higher on some criterion, people will first attempt to use criterion knowledge (direct knowledge of the answer) or elementary logic to inform their decisions. If no such criterion knowledge or logic existed, they proposed people would activate from declarative knowledge a probabilistic mental model. Such a model would consist of probabilistic cues, in other words, facts about an object that are correlated with the criterion. Subjective recognition of an item was one of these cues, which Gigerenzer et al. referred to as the “familiarity cue”.

In a subsequent article, Gigerenzer and Goldstein (1996) asserted that recognition holds a special status among probabilistic cues, because if an object is not recognized, it is not possible
to recall cue values for that object from memory. In this sense, they claimed that recognition preceded cue recall and served as an initial “screening step” that occurred prior to searching for further cue information. Re-interpreted from a dual-process perspective, their use of “recognition” is analogous to familiarity, whereas their use of “recall” is consistent with recollection. Based on these early assertions by Goldstein & Gigerenzer, one might suggest that RH decisions are based on an initial sense of familiarity, a “screening step” that precedes recollection of other cues or knowledge. If two items can be dissociated based solely upon their early respective familiarities (as they should if one item is recognized and the other is completely novel), it would be unnecessary to probe memory for further cues. This strategy is in line with the common “less is more” mantra advocated by the fast and frugal heuristics program, as people would capitalize on an early sense of familiarity to make their decisions, and forego any further effort to recollect relevant information.

Upon their formalization of the RH, Goldstein and Gigerenzer (2002) reconstituted the assertion of recognition as a noncompensatory cue guiding RH-based decisions. That is, they explicitly stated that if one object is recognized but not the other, all other cue knowledge pertaining to the recognized object is ignored and an inference is based exclusively on the recognition cue. This noncompensatory claim lends further credence to the idea that an early familiarity process guides RH-based decisions, insofar as recollection of any kind is deemed irrelevant. The boldness of this assertion has been challenged by other groups, several of whom found that choices between pairs of items for which only one item is recognized and some additional knowledge is available are better predicted by the RH than pairs of items where only one item is recognized but no additional knowledge is available (Newell & Fernandez, 2006; Pohl, 2006). However, these experiments failed to demonstrate that participants were actively
using this additional knowledge when making decisions, and therefore it is possible that items associated with additional knowledge were also associated with a greater sense of familiarity. This heightened sense of familiarity could then support the finding of greater adherence to the RH in cases where additional knowledge was available. Taken altogether, previous research surrounding the RH has predominantly promoted familiarity as the primary contributor to recognition-based decisions, though few studies have formally placed familiarity within a dual-process account of memory when considering its role.

There is an extensive amount of research demonstrating that event-related potentials (ERPs) are able to dissociate the dual-process contribution of familiarity and recollection to recognition memory (Curran, 2000; Friedman & Johnson, 2000; Opitz & Cornell, 2006; Rugg & Curran, 2007). Two ERPs that are both temporally and topographically distinct have been specifically associated with familiarity and recollection. Familiar stimuli elicit more positive-going ERP waveforms than unfamiliar stimuli at fronto-central recording sites between 300-500 ms, an effect commonly referred to as the ‘FN400’ (e.g., Curran, 2000). Recollection is associated with a parietal maximally-positive ERP that onsets around 450 ms post-stimulus until around 800 ms, and has been termed simply the “parietal old/new effect” (e.g., Jäger, Mecklinger, & Kipp, 2006).

Rosburg, Mecklinger, and Frings (2011) endorsed a dual-process familiarity-based approach to the RH that implemented a city-size comparison task while recording ERPs. Cities with previously established recognition rates were paired so that well-known cities (recognition rates > 80%) were always paired with little-known cities (recognition rates < 15%). ERPs were recorded for each city presentation within a pair, and subsequently all well-known cities were averaged together and all little-known cities were averaged together. Their results showed
pronounced differences for ERPs in response to well-known and little-known city names during the 300-450 ms window (roughly corresponding to the FN400) as well as the 450-600 ms window (roughly corresponding to the parietal old/new effect). From 300-450 ms, well-known city names elicited more positive-going waveforms, particularly at fronto-central electrode sites, consistent with greater familiarity for well-known cities. From 450-600 ms, well-known city names continued to elicit more positive-going waveforms, this time particularly at parietal sites, consistent with greater recollection for well-known cities. Additionally, pattern classification techniques were implemented to predict participants’ decisions on a trial-by-trial basis using data from the FN400 and parietal old/new effect time windows, suggesting that these processes were having a real impact on actual choices.

Rosburg et al.’s (2011) dialogue highlights the significance of FN400 familiarity effects in dissociating recognized from unrecognized cities and their potential usefulness in RH-based decisions, but leave some skepticism surrounding the also significant parietal old/new recollection effects. They trained pattern classification models that accurately predicted participants’ decisions based on single trial EEG data. An initial model included only the FN400 time window, and showed that the classifier accurately predicted participants’ decisions based on this time window alone. A second model incorporated data from the parietal old/new effects time window, and showed a slight increase in classifier accuracy from the previous model. However, a model consisting solely of parietal old/new effects was not tested, and thus the role of recollection in RH-based decisions is more difficult to ascertain from this work. The authors concede that recollection of some sort is likely occurring, as indicated by significantly larger parietal old/new effects for well-known cities, and posit that participants could be recollecting any kind of available information associated with the given city name. Although the existence of
differential FN400s (indexing familiarity) for well-known and little-known cities would be unanimously predicted during the city-size comparison task, the existence of differential parietal old/new effects (indexing recollection) during this decision-making task might be more controversial from an RH perspective. If we believe the RH is a truly noncompensatory mechanism, and decisions are made based solely on mere recognition of one item, we might expect all the information necessary to make a decision to be located in familiarity differences. However, it is not unlikely that cities that are familiar are also associated with a certain amount of knowledge in memory. This knowledge might be automatically recollected upon presentation of the city’s name, and could account for the parietal old/new effect differences reported by Rosburg et al. Alternatively, a strict interpretation of the RH might suggest that because all the information necessary to make a decision is available based on familiarity differences, participants acutely tuned to the RH should abandon any further recollection beyond familiarity, and therefore parietal old/new effects would not be expected to differ for well-known and little-known cities during the decision making task. Regardless of which perspective of recollection within the RH is adopted, the noncompensatory component of the RH is not threatened by Rosburg et al.’s finding of significant parietal old/new effects unless participants are actually utilizing recollected knowledge to make decisions.

In summary, multiple theoretical accounts in the literature and the empirical ERP findings reported by Rosburg et al. (2011) both point to a role for familiarity in RH-based decisions. In environments where recognition is an ecologically rational cue, or recognition is correlated with a given criterion (e.g. city population), a sense of familiarity should help guide recognition-based decisions. Less theoretical emphasis is placed on a role for recollection, due in
part to the noncompensatory nature of the RH, which asserts that any recollected knowledge should not be considered in RH-based decisions.

**The Fluency Heuristic**

Research surrounding the FH has also been limited with respect to directly addressing potential contributions of familiarity and recollection. However, similar to the RH, several theoretical claims seem to point to familiarity as the main contributor to FH-based decisions. Before considering these claims, it is necessary to specifically define “fluency” as it relates to the FH. Many variants of fluency exist in the literature, including, for example, absolute, relative, conceptual, and perceptual. The FH, as formalized by Schooler and Hertwig (2005), refers specifically to the speed at which a stimulus can be processed or retrieved from memory. Faster recognized items are considered more fluent, and people attribute fluent processing of stimuli to having experienced the stimuli before. More frequent and meaningful exposure to a stimulus in the environment is said to lead to more fluent processing. Several early fluency researchers have demonstrated this concept. For example, researchers have tampered with the previous exposure of certain stimuli to increase the perceived fame of nonfamous names (the false fame effect; Jacoby et al., 1989) and the perceived truth of repeated assertions (the reiteration effect; Begg et al., 1992; Hertwig et al., 1997). These researchers predominantly suggested that increasing exposure to a given stimulus increases its familiarity, and thereby its fluency. More recently, however, Kurilla and Westerman (2008) conducted a study that demonstrated that enhancing perceptual and conceptual fluency reliably increased claims of both familiarity and recollection. So, fluency has been shown to influence perceived memory judgments across a plethora of domains. A problem with this general line of research remains the slippery nature of the word ‘fluency’, and researchers’ tendencies to interpret it differently across studies. Fluency has been
referred to as “the subjective experience of ease” (Oppenheimer, 2008), “the subjective feeling of familiarity” (Kelley & Jacoby, 1998), and “easy or efficient processing” (Whittlesea & Leboe, 2003), among others. These definitions are often imprecise, which has led to difficulty objectively measuring fluency.

Schooler and Hertwig’s (2005) formalization of fluency and the FH was theoretical in nature. They implemented the FH and the RH within the ACT-R cognitive architecture (Anderson et al., 2004), and were therefore able to precisely define fluency in terms of the time it takes to retrieve a memory trace (or chunk, to use ACT-R terminology). This specific definition was coined retrieval fluency, and Schooler and Hertwig asserted that the FH was able to tap indirectly, via retrieval fluency, into the environmental frequency information locked in the chunks’ activation values. Marewski and Mehlhorn (2011) furthered the work integrating the RH and FH within the ACT-R architecture, and importantly, their instantiation of the models assumed that people would first assess recognition of city names before potentially attempting to retrieve any further cues. Their use of “recognition” is taken to be synonymous with familiarity, and they cite dual-process literature supporting the idea that familiarity arrives on the mental stage earlier than recollection. So in this sense, familiarity is assessed first in the model before attempting any type of recollection. This ACT-R interpretation of the RH and FH implies a necessary role for familiarity, and leaves the door open for a potential role for recollection.

Hertwig, Herzog, Schooler, and Reimer (2008) took Schooler and Hertwig’s (2005) research a step further, using empirical data to test several theoretical claims regarding the FH. Among their most significant findings, they showed that (a) retrieval fluency correlates with objective properties of the world, such as a city’s population (i.e. retrieval fluency is an ecologically rational cue), and (b) that people can accurately introspect to discriminate between
differences in their own retrieval fluencies. They also showed that people’s decisions adhered to the FH more frequently when a difference between retrieval fluencies for two items was large, where recognition latency (the reaction time it takes for a subject to identify an item as recognized) was used as a proxy for retrieval fluency. This finding suggests that the more disparate retrieval fluencies are between items, the more useful the FH is. In their review of previous literature, Hertwig et al. abstract across different meanings of the FH and conclude that the gist involves three properties: (a) the attribution of “fluent”, or speedy, processing to prior experience, (b) the resulting conscious experience of familiarity, and (c) the assumption that relative fluency between a pair of items can be used as a basis for recognition memory. Their assertion of a conscious assessment of familiarity during FH-based decisions firmly identifies a role for familiarity in the FH, with no mention of a possible role for recollection. Importantly, Hertwig et al.’s main goal in their 2008 paper was to enforce the idea that inferences could be made, and were indeed made, based on retrieval fluency differences for a pair of objects in a single-cue and noncompensatory fashion. So, to the extent that fluency might reference different levels of familiarity, it could be argued that the FH relies indirectly on a familiarity distinction between two objects.

Similar to the RH, Schooler and Hertwig (2005; see also Hertwig et al. 2008) asserted that FH-based inferences were made using retrieval fluency in a noncompensatory fashion. Though the authors certainly demonstrated that fluency affects judgments, none of their experiments directly showed that participants relied on a fluency cue in isolation when making inferences. These same arguments can be made against the noncompensatory claim of the RH. The vast majority of research on both heuristics has relied on adherence rates, or accordance rates, to quantify usage. Adherence rates are calculated by taking the number of trials a
participant’s responses are in line with a certain heuristic (i.e., for the RH, actually choosing the recognized city as being more populous), and dividing that by the total number of trials in which that heuristic was applicable. This calculation results in a biased, though not inconsequential approximation of a given heuristic’s use. Adherence rates are biased because observed choices in line with a heuristic’s prediction cannot imply that this heuristic was actually used (e.g., Fielder, 2010; Hilbig & Pohl, 2008). In both the case of the RH and the FH, further knowledge or information that ultimately argued for the chosen object may well have been considered. For cases where only one object is recognized, any further knowledge available about the recognized object is confounded with its mere recognition during decisions, and adherence rates cannot identify which source is used in making a choice. For cases where both objects are recognized, further knowledge is confounded with retrieval fluency, and it is probable that objects recognized more fluently are also associated with more readily accessible knowledge that is also more valid. So again, adherence rates cannot capture which source of information is contributing to decisions.

Researchers worked around this obstacle until Hilbig, Erdfelder, and Pohl (2010) created a multinomial processing tree (MPT) model that was able to provide an unbiased measure of RH use. In general, MPT models make use of categorical data to estimate latent probabilities of various cognitive processes underlying task performance, and thus can provide insight into the contribution of different processes to an observed outcome (Batchelder & Riefer, 1999). After the initial success and validation of their first model, Hilbig, Erdfelder, and Pohl (2011) extended the model to incorporate an unbiased measure of FH use in addition to RH use; they coined this model the r-s-model (see Figure 1).
**Fig. 1.** Processing tree representation of the r-s-model, re-printed with permission from Hilbig et al. (2011). Parameters include: Recognition validity ($a$), fluency validity ($c$), knowledge validities ($b_1$, $b_2$, $b_3$), probability of valid guesses ($g$), probabilities of using the recognition heuristic (RH; $r$), and probabilities of using the fluency heuristic (FH; $s$). Boxes with rounded corners signify latent states.
Figure 1 shows the r-s-model in its entirety, which consists of four separate trees representing four possible cases: a) both objects are recognized, and their respective difference in retrieval fluencies is less than 100 ms (“fluency-homogeneous knowledge cases”), b) both objects are recognized, and their respective difference in retrieval fluencies is greater than 100 ms (“fluency-heterogeneous knowledge cases”, corresponding to FH trials), c) only one object is recognized (“recognition cases”, corresponding to RH trials), or d) neither object is recognized (“guessing cases”). Knowledge cases are divided into fluency-homogeneous and fluency-heterogeneous conditions in accordance with Hertwig et al.’s (2008) finding that retrieval fluency differences below a 100 ms threshold were indistinguishable by participants, and therefore the fluency cue was unavailable in these conditions. Empirically observed judgments for each of these four cases are further categorized as correct or false with respect to the true criterion (e.g., city population), as well as whether a given choice adhered to the applicable heuristic (i.e., the recognized object was chosen during RH cases, and the more fluently retrieved object was chosen during FH cases). Taken together, these categories combine for a total of 12 observable outcomes (labeled in the right column of Figure 1), 9 of which are free, that are explained through 8 free estimated parameters. Consequently, the $\chi^2$-goodness-of-fit test of the r-s-model has $9 - 8 = 1$ degree of freedom. Using this information, it is possible to first test the fit of the model statistically, and then obtain more useful model parameter estimates. The two parameters of greatest interest are the $r$-parameter and $s$-parameter (for speed-based judgments), which indicate use of the RH and FH, respectively.

Because the $r$-parameter and $s$-parameter incorporate extra categorical information that regular adherence rates do not, it is possible for these estimates to unconfound the actual heuristic cue in question (i.e., recognition or fluency) from further knowledge that might have
been used in the decision. Previous versions of the r-s-model (that focused only on RH cases) typically found that observed measures of RH use were typically approximately 15-20% lower than was indicated by traditional adherence rates, with an average $r$-parameter of 63% (Hilbig, 2010; Hilbig et al., 2010). Importantly, most data sets tested showed that participants still used the RH at an above-chance level. However, the emergence of the r-s-model as a valid measurement tool for FH decisions indicated a more precipitous drop-off in observed FH use; down from common adherence rates around 60-70% to just 21% as estimated by the $s$-parameter (Hilbig et al., 2011). This estimate dropped FH use below chance, and suggested that participants were only capitalizing on retrieval fluency information in about one-fifth of the trials where it was applicable. Additionally, this figure suggests that in approximately 40-50% of those same trials where participants’ decisions were in line with the FH, but did not rely on retrieval fluency in isolation, people resorted to an alternative strategy. Because more fluent cities are typically associated with more prior experience in the world, it seems logical that participants’ choices would coincide with the more fluent city as being larger more often than not. However, the question remains whether familiarity, recollection, or a combination of both processes could be playing a role in these decisions. As previously mentioned, familiarity has been discussed as a mechanism potentially contributing to FH decisions, insofar as familiarity contributes to the retrieval fluency of different objects. However, if recollection processes were being utilized in decisions, this would indicate that participants were taking advantage of further knowledge available about a stimulus. Use of this further knowledge to make decisions would not only challenge the noncompensatory claim of the FH, but would suggest that perhaps reliance on further knowledge is a more useful cue than retrieval fluency, and challenge the viability of the FH as a model of comparative judgments.
In summary, attempts to directly assess dual-process contributions of familiarity and recollection in memory-based heuristics are few and far between, with the most direct attempt coming from Rosburg et al.’s (2011) evaluation of familiarity’s role in the RH. Experiment 1 aims to use recognition memory ERPs to investigate the unique contributions of familiarity and recollection to the RH and FH. We implemented an adapted version of the standard city-size comparison task. Instead of pairing cities based on a priori recognition rates, as was done by Rosburg et al. (2011), we coupled the inference task with a recognition test in order to obtain recognition responses unique to each subject for each city included in the experiment. This design allows us to examine situations where either the RH or FH was applicable, based on individuals’ own recognition responses.

Implementing this design, we expect inference trials where the RH was applicable (participants judged one city as recognized but not the other) to roughly replicate Rosburg et al.’s (2011) results for well-known cities vs. little known cities. We anticipate more positive FN400s, consistent with greater familiarity, for cities in the inference task judged as recognized compared to unrecognized. If RH-based decisions were truly utilizing familiarity in a noncompensatory fashion, we would not expect parietal old/new effects (indexing recollection) to significantly differ between recognized and unrecognized stimuli within a pair. Rosburg et al. found more positive parietal old/new effects for recognized compared to unrecognized cities, but differences in our task methodology and analysis lead us to predict that parietal old/new effects will not differ for recognized vs. unrecognized cities. In particular, our task implements a recognition test, allowing us to purely dissociate RH trials from FH trials, unlike Rosburg et al.’s task, which assumed all trials were RH trials based on a-priori recognition rates. Additionally, Rosburg et al.
chose an earlier and shorter time window for parietal old/new effects (450 - 600 ms) than we will analyze (500 - 800 ms).

For FH trials, retrieval fluencies (the time it took to identify a city as recognized) for each city within a pair were used to quantify individual trials as a large recognition latency difference (>400 ms difference between the two cities) or a small recognition latency difference (<400 ms difference between the two cities). This 400 ms cutoff was chosen in accordance with previous studies on the FH (Hertwig et al., 2008; Volz, Schooler, & Cramon, 2010). City pairs with a small latency difference between the two should be similar to each other in terms of familiarity and recollection, as indexed by similar FN400 and parietal old/new effects, and fluency should not be a useful cue in these scenarios. We expect city pairs with a large latency difference between the two to have divergent ERPs. Logically, a city that takes longer to be recognized might be less familiar and/or more poorly recollected compared to a city with a shorter recognition latency (though this claim is evaluated in Experiment 2). If familiarity is truly driving FH-based decisions, as the literature has implied, we would predict more positive FN400s for cities with shorter recognition latencies within a pair, consistent with greater familiarity for these cities. Additionally, if recollection is playing a role in these decisions via the retrieval of cue knowledge, we would expect more positive parietal old/new effects for cities with shorter recognition latencies within a pair, consistent with greater recollection for these cities.

Experiment 1

Method

Participants
Fifty-nine right-handed participants ranging in age from 18-29 took part in the study. Data from 11 participants were excluded due to technical artifacts or low trial counts (less than 15 trials per condition). Of the remaining 48 subjects used for analyses, 21 were female. Recruited participants were either paid volunteers ($15/hour) or undergraduate students receiving course credit from the University of Colorado. All participants were informed about the procedure and gave their written consent before participating.

**Materials and Procedure**

Each participant performed two computerized tasks while EEG data were recorded: a city/country recognition test and a population inference task. Task order was counterbalanced across subjects, as was order of presentation of cities and countries. Prior to beginning the experiment, each participant completed a 1-min practice session for the population inference task. Stimuli were the same for both tasks: U.S. city and country names displayed in the center of a computer monitor, one at a time. The stimuli that appeared in the practice session did not appear in the actual task.

For the recognition test, participants viewed the 100 most populous cities in the United States and the 100 most populous countries in the world, as well as 10 fictional cities and 10 fictional countries intended to increase the honesty of responses. City and country names (hereby referred to collectively as ‘regions’) were displayed on the screen one at a time in random order, with separate counterbalanced city and country blocks. Participants were instructed to indicate with a button press, as quickly and accurately as possible, whether or not they were familiar with each region from prior to the experiment. Stimuli remained on the screen for a minimum of 2 s or until a response was made. If a response was not made within 4 s, a question mark prompt (“?”) appeared, encouraging participants to respond. A 1000 ms interstimulus fixation-cross
followed each response before the next region was presented. Left and right response
(“yes”/“no”) key assignments were counterbalanced. Participants were instructed to use their left
and right index fingers to respond. Reaction times (interpreted as recognition latencies) and
responses were recorded with response boxes accurate to within 1 ms.

The population inference task closely mirrored the design of Rosburg et al. (2011), and
each participant performed the same task for two conditions: U.S. cities and countries (see Figure
2 for a sample trial sequence). Order of task condition was counterbalanced across participants.
Stimuli in the inference task are identical to those in the recognition test, with the exception of
the fictional region names, which were excluded from the inference task. Each trial consists of
four screens, which are all separated by a centered fixation-cross for 1000 ms. The first screen
asks participants to concentrate for 1000 ms. Next, the first region name is then centrally
displayed for 2000 ms, followed by the second region name for the same duration of 2000 ms.
Lastly, participants viewed the decision frame where the two regions are re-presented, and were
instructed to select the region they believed had more inhabitants. Region 1 is always located
above Region 2 in the decision frame, with a fixation-cross centrally located between them. For
the inference task the response box was situated vertically, with the uppermost button
corresponding to Region 1, and the lowermost button corresponding to Region 2. Half the
participants were instructed to place their left index finger on the uppermost button and right
index finger on the lowermost button, and the other half vice versa. The final decision frame
remains on the screen until a response is made, at which point a new trial begins with the “Please
Concentrate” screen.
In each condition of the inference task, there were eight blocks of 25 trials for a total of 200 inferences per condition. Subject-timed breaks occurred after every 8 trials. Stimuli were drawn randomly without replacement for the first two blocks (50 inferences), so that every region was shown once. This process repeated for a total of eight blocks for each condition, so that participants viewed each region exactly four times. In rare cases the same two regions were paired more than once for the same participant, though not frequently enough to impact the results. In between city and country conditions, participants were allowed to rest before completing the second half of the inference task. Responses were collected while continuous EEG was recorded throughout the task.

EEG/ERP Methods

EEG (electroencephalogram) was collected with a 128-channel HydroCel Geodesic Sensor Net connected to AC-coupled, 128-channel, high-input impedance amplifiers (Electrical Geodesics Inc., Eugene, OR). Amplified voltages were digitized at 250 Hz. Individual sensors were adjusted at ~20-min intervals until impedances were less than 50 kΩ.

The EEG was digitally low-pass filtered at 40 Hz and high-pass filtered at .1 Hz prior to ERP analysis. Trials were discarded from analyses if they contained eye movements (vertical
EOG channel differences greater than 70 µV) or more than 20 bad channels (changing more than 100 µV between samples, or reaching amplitudes over 200 µV). Individual bad channels in trials with less than 20 total bad channels were replaced on a trial-by-trial basis with a spherical spline algorithm (Srinivasan, Nunez, Silberstein, Tucker, & Cadusch, 1996). EEG was collected with respect to a vertex reference, and ERPs were re-referenced to an average reference. ERPs were baseline corrected to a 200 ms pre-stimulus recording interval.

Results

Behavioral

During the recognition test, participants identified an average of 156 regions as recognized and 64 as not recognized. This resulted in an average of 296 inference trials/subject where both regions within a pair were recognized (FH trials) and 93 inference trials/subject where only one region within a pair was recognized (RH trials). Additionally, participants on average only claimed to recognize 1.3 out of the 20 fictional regions included in the recognition test, so responses were honest. Because our stimulus set was slightly different than what has traditionally been used in city-size comparison tasks, we first assessed the operational statistics to ensure comparability with previous studies. The most important factor for determining the useful of these two heuristics within a given domain is the strength of the relationship between memory and the criterion of interest (Goldstein & Gigerenzer, 2002; Schooler & Hertwig, 2005), which in this case is population of U.S. cities and countries. To assess this strength we can calculate what has been termed the recognition validity for RH cases and the fluency validity for FH cases. The recognition validity is defined as the proportion of RH trials where the recognized region is actually more populous than the unrecognized region, regardless of the participant’s decision. Similarly, the fluency validity is defined as the proportion of FH trials where the more
quickly retrieved region is actually more populous than the less quickly retrieved region, regardless of the participant’s decision. This number also reflects what the participants’ highest attainable accuracy would be in the task if they adhered to the heuristic on all trials, and in this way gives a sense of the “ecological rationality” of a given heuristic within a certain domain (Goldstein & Gigerenzer, 2002).

For our stimulus set, the recognition validity ($M = .76$, $SD = .06$) and fluency validity ($M = .57$, $SD = .04$) were within range of previously reported findings (Hertwig et al., 2008; Hilbig et al., 2011), and both were significantly greater than chance ($t(47) = 28.1$, $p < .0001$; $t(47) = 13.0$, $p < .0001$; respectively) indicating that recognition was an ecologically rational cue during RH trials, and fluency was an ecologically rational cue during FH trials.

In order to assess how frequently participants’ actual choices were in line with each heuristic’s predictions, adherence rates were calculated. The RH adherence rate ($a_r$) was calculated:

$$a_r = \frac{R_r}{R_r + NR_r}$$

where $R_r$ are RH trials where the participant selected the recognized region within a pair as being larger and $NR_r$ are RH trials where the participant selected the unrecognized region within a pair as being larger. RH adherence was 89.4%, indicating participants’ inferences adhered to the RH at an above-chance level ($t(47)=37.21$, $p<.0001$). Likewise, the FH adherence rate can be calculated:

$$a_f = \frac{F_f}{F_f + S_f}$$

where $F_f$ are FH trials where the participant chose the region with a faster reaction time within a pair as being larger and $S_f$ are FH trials where the participant chose the region with a slower reaction time within a pair as being larger. The overall FH adherence rate was 59%, significantly
above chance \((t(47)=13.06, p<.0001)\). FH adherence rates were also calculated separately for trials with large recognition latency differences (>400 ms) and trials with small recognition latency differences (<400 ms). For trials with large differences, FH adherence was 66.2%, significantly above chance \((t(47)=13.62, p<.0001)\). For trials with small differences, FH adherence was 55.9%, also significantly above chance \((t(47)=8.23, p<.0001)\). Additionally, FH adherence rates were significantly greater for trials with large recognition latency differences compared to trials with small differences \((t(94)=7.41, p < .0001)\).

Accuracy across all inference trials was 68.1%, indicating participants correctly chose the more populous region at a rate significantly above chance \((t(47)=38.79, p < .0001)\). Accuracy was also calculated separately for adherent and non-adherent trials. For RH-adherent trials, where participants chose the recognized region as being larger, accuracy was 77.4%, significantly above chance \((t(47)=18.01, p < .0001)\). For non-adherent RH trials, where participants chose the unrecognized region as being larger, accuracy was 35.8%, significantly below chance \((t(47)=10.91, p < .0001)\). For FH-adherent trials, where the participant chose the more quickly retrieved region as being larger, accuracy was 70.5%, significantly above chance \((t(47)=16.32, p < .0001)\). For non-adherent FH trials, where the participant chose the less quickly retrieved region as being larger, accuracy was 61.9%, significantly above chance \((t(47)=10.07, p < .0001)\), though also significantly lower than adherent trials \((M = 70.5, t(92)=7.18, p < .0001)\). Overall, participants performed well on the inference task and were typically more accurate when their responses adhered to the heuristics compared to when they did not.

**Multinomial Processing Tree analysis**

As noted previously, it should be emphasized that although participants’ choices may be in line with a given heuristic’s prediction, simple adherence rates alone cannot imply that this
heuristic was actually used (e.g. Fielder, 2010; Hilbig & Pohl, 2008). Hilbig et al.’s (2011) r-s-model utilizes participants’ recognition judgments, corresponding recognition latencies, and actual choices in the judgment task to categorize each inference trial into one of 12 observable outcomes, allowing for a more complete picture of noncompensatory RH- and FH-use. To apply the r-s-model, we first computed the frequency of the 12 observable outcomes across all participants (see Appendix A) and then used standard software for multinomial processing tree (MPT) modeling (Moshagen, 2010) to obtain parameter estimates and the overall fit of the r-s-model. Forty-eight participants resulted in an aggregate of 19,200 inference trials. Considering the large number of trials and high statistical power for a goodness of fit test, the model fit the data well ($G^2(1) = 3.57, p = .06$)\(^1\). It should be noted that for the log-likelihood ratio statistic $G^2$ larger $p$-values indicate a better model fit. With over 19,000 observations, the $G^2$ test has substantially greater power than is typical in categorical frequency data, and even miniscule deviations from the perfect model fit are very likely to be detected. In turn, $p > .05$ should be considered a superior fit. Parameter estimates are displayed in Table 1.

\(^1\) The r-s-model was additionally fit separately for city stimuli ($G^2(1) = 0.25, p = .62$) and country stimuli $G^2(1) = 6.28, p = .01$. While the city stimuli fit the overall model better than the country stimuli, both resulted in a highly similar pattern of parameter estimates, suggesting similar underlying processes for both sets of stimuli.
The model-estimated recognition validity \((M = .76)\) was identical to that reported in the observational statistics above \((M = .76)\), and the model-estimated fluency validity \((M = .59)\) was nearly identical to that reported in the observational statistics above \((M = .57)\). The similarity of these validities corroborates the overall validity of the model. The two parameter estimates of greatest importance are the probability of RH-use based on recognition alone \((r\)-parameter\) and the probability of FH-use based on retrieval fluency alone \((s\)-parameter\). According to the \(r\)-\(s\)-model, participants relied on the recognition cue in isolation when one region was recognized and the other was not recognized on 76% of the trials. This estimate is lower than the mean adherence rate reported above \((M = .89)\), though still significantly greater than chance \((\Delta G^2 = 357, p < .0001, \text{when fixing } r = .50)\), indicating that participants relied on the RH on a substantial portion of trials.

For FH trials, the \(r\)-\(s\)-model estimated that participants relied on retrieval fluency in isolation when both regions were recognized on only 16% of the trials. This \(s\)-parameter closely replicates Hilbig et al.’s (2011) findings, indicating that participants only used the FH on approximately one fifth of the trials in which it could have been applied. This estimate \((s = .16)\),

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Psychological Meaning</th>
<th>Estimate ((SE))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
<td>recognition validity</td>
<td>.76 (.01)</td>
</tr>
<tr>
<td>(b1)</td>
<td>knowledge validity, fluency-homogenous FH cases</td>
<td>.66 (.01)</td>
</tr>
<tr>
<td>(b2)</td>
<td>knowledge validity, fluency-heterogeneous FH cases</td>
<td>.68 (.01)</td>
</tr>
<tr>
<td>(b3)</td>
<td>knowledge validity, RH cases</td>
<td>.67*</td>
</tr>
<tr>
<td>(c)</td>
<td>fluency validity</td>
<td>.59 (.00)</td>
</tr>
<tr>
<td>(g)</td>
<td>correct guessing (neither object is recognized)</td>
<td>.56 (.02)</td>
</tr>
<tr>
<td>(p)</td>
<td>proportion of fluency-homogenous FH cases</td>
<td>.30 (.00)</td>
</tr>
<tr>
<td>(r)</td>
<td>RH-use (considering the recognition cue in isolation)</td>
<td>.76 (.01)</td>
</tr>
<tr>
<td>(s)</td>
<td>FH-use (considering retrieval fluency in isolation)</td>
<td>.16 (.01)</td>
</tr>
</tbody>
</table>

Table 1. Parameters of the \(r\)-\(s\)-model, psychological meaning of the parameters, and parameter estimates with standard errors of each estimate, based on data from all participants.

* This number is derived analytically from \(b_3 = p x b_1 + (1 - p) x b_2\) and is thus reported without a standard error.
which unconfounds the contributions of retrieval fluency and further knowledge or other information, is much lower than the mean FH adherence rates for trials with small RT differences ($M = .56$) as well as trials with large RT differences ($M = .66$). By the nature of the r-s-model, it is not possible to calculate s-parameters separately for trials with small RT differences (0-400ms) and trials with large RT differences (400+ms). This is because the model assumes a fluency threshold of 100 ms, such that all trials with a RT difference less than 100 ms are considered to be indistinguishable based on retrieval fluency, and these observations are placed in a separate branch of the model. Selectively excluding trials based on RT differences would invalidate the model. However, Hilbig et al. (2011) experimented with implementing different fluency thresholds in the model, beginning at 0 ms and increasing in steps of 100 ms to 1,000 ms, and observed the effect on the s-parameter. They found that the s-parameter is robust across fluency thresholds, with adjustment of the threshold failing to increase or decrease the s-parameter estimate by more than 11%. Because Hilbig et al.’s (2011) data and our own fit the same underlying model, it is fair to assume that our s-parameters would be unlikely to differ by more than 11% for trials with small RT differences and trials with large RT differences.

However, it is still valid to compare the overall FH adherence rate ($M = 59\%$) to the s-parameter estimate output by the r-s-model ($s = 16\%$), and we find that the proportion of cases in which participants relied on retrieval fluency in isolation is significantly lower than the mean FH adherence rate ($\Delta G^2 = 2675, p < .0001$, when fixing $s = .59$) So, according to the r-s-model, pure use of the FH seems to be universally less common than originally thought, regardless of reaction time difference. We will return to this methodology in more detail in the Experiment 2 results section.
ERP Analyses

Based on previous studies that have examined the ERP correlates of familiarity and recollection, we selected a priori time windows of 300-500 ms and 500-800 ms post stimulus onset (Hayama, Johnson, & Rugg, 2008; Rugg & Curran, 2007; Woodruff et al., 2006; among others). In each time window, mean amplitudes were extracted and a 3 (condition: RH, FH < 400, FH > 400) x 2 (recognizability\(^2\): more recognizable, less recognizable) x 2 (posteriority: anterior clusters, posterior clusters) x 2 (laterality: left-hemisphere clusters, right-hemisphere clusters) repeated measures ANOVA was conducted. Task order (recognition test or inference task first) was included as a between subjects variable, but did not reach significance in any analyses. Condition was broken down to RH trials, FH trials with small recognition latency differences (<400 ms), and FH trials with large recognition latency differences (>400 ms). The 400 ms cutoff was selected based on previous research (Hertwig et al., 2008; Volz et al., 2010). For RH trials, recognized regions were considered ‘more recognizable’ and unrecognized regions were considered ‘less recognizable’. For FH trials, regions with longer recognition latencies within a pair were considered ‘more recognizable’, and those with shorter recognitions latencies were considered ‘less recognizable’.

Four regions of interest (ROIs) were selected for analysis based on those used in other studies (e.g., Curran, 2004; Curran, DeBuse, & Leynes, 2007; Curran & Friedman, 2004; Mollison & Curran, 2012), each composed of an average of 7 electrodes (see Figure 3 below). The ROIs were labeled as follows: LAS: Left anterior-superior, RAS: Right anterior-superior, LPS: Left posterior-superior, RPS: Right posterior-superior. These four clusters enabled us to look at both anterior/posterior distinctions and hemispheric laterality distinctions apparent in the

\(^2\) This variable refers to the more recognized and less recognized city within a pair (trial), based on behavioral data from the recognition test. We had trouble naming this variable because it refers to two different types of stimuli for RH trials and FH trials. This distinction is described within the text.
ERP waveforms at different time windows. Figure 4 (see below) shows plots of grand average ERPs of the four ROIs for the three conditions (RH, FH < 400, FH > 400). Figure 5 (see below) shows corresponding topographic plots of the entire scalp for all three conditions (RH, FH < 400, FH > 400) at the early (300-500 ms) and late (500-800 ms) time windows. Figure 6 shows a bar graph of corresponding mean ERP amplitude differences.

Fig. 3. The 128-channel HydroCel Geodesic Sensor Net™ used to measure the EEG and regions of interest (ROIs) on which the analysis were based. Each ROI label describes its position on the skull: R, right; L, left; A, anterior; P, posterior. ERPs displayed are averaged across anterior sites (AS) and posterior sites (PS).
**Fig. 4** Event-related potentials (ERPs) averaged across responses for recognition heuristic (RH) trials, Fluency Heuristic (FH) trials with small recognition latency (RT) differences, and FH trials with large RT differences. ERP waveforms are shown from -200 ms pre-stimulus to 1000 ms post-stimulus. Positive is plotted upward. The two time windows of interest (300-500ms, 500-800ms) are indicated by dotted vertical lines. Asterisks denote significant differences in ERP mean amplitudes for the corresponding time window.

**Fig. 5** Topographic maps of voltage amplitude differences across the entire scalp. The left column shows recognition heuristic trials (RH): activation for the unrecognized region is subtracted from recognized regions (R-NR) for the early time window associated with familiarity (300-500 ms) and the late time window associated with recollection (500-800 ms). The middle column shows activation differences for fluency heuristic trials with a small recognition latency difference (FH Small): the region with a longer recognition latency within a pair subtracted from the region with a shorter recognition latency within a pair. The right column shows activation differences for fluency heuristic trials with a large recognition latency difference (FH Large): the region with a longer recognition latency within a pair subtracted from the region with a shorter recognition latency within a pair. Red regions indicate greater amplitudes for well-known regions within a pair, and blue regions indicate greater amplitudes for the lesser-known regions within a pair.
In order to overcome the potentially deleterious effects of variance heterogeneity common in psychophysiological studies, we adopted a non-parametric statistical procedure based on robust estimators instead of the usual repeated measures ANOVA based on least squares parameters. This approach implements an approximate degrees of freedom test statistic based on Winsorized variances that have been shown to circumvent the biasing effects of variance heterogeneity (Keselman, Algina, Lix, Wilcox, & Deering, 2008). Only significant and/or relevant main effects and interactions are reported.

Trials in which participants’ decisions adhered to the heuristics were analyzed separately from trials in which participants’ decisions were not adherent to the predictions of each heuristic.
However, due to exceedingly low trial counts for non-adherent RH trials, ERPs could not be reliably computed for this condition. Again due to low trial counts for the remaining two conditions (non-adherent FH > 400 and non-adherent FH < 400), 13 additional subjects were excluded from analysis. Non-adherent analysis was conducted on a subset of 35 subjects with at least 15 trials in each of the two remaining conditions, as opposed to the 48 subjects included for adherent analyses. Similar ANOVAs were run on adherent and non-adherent trial data, with the exception of the condition variable having three levels (RH, FH < 400, FH > 400) for adherent analyses and only two levels (FH < 400, FH > 400) for non-adherent analyses.

300 - 500 ms

Adherent Trials

The ANOVA for the time window corresponding to familiarity (300 – 500 ms) revealed a significant three-way interaction between condition, recognizability, and posteriority ($T_{W3}/c(2.0,38.7)=3.69, p=.034$), such that there was only a two-way condition by recognizability interaction present for the anterior ROIs ($T_{W3}/c(2.0,37.6)=5.03, p=.018$). Within the anterior ROIs, only the RH condition yielded a significant effect of recognizability, such that ERPs in response to unrecognized regions within a pair were significantly more negative than those in response to recognized regions ($T_{W3}/c(1.0,41.4)=8.73, p=.005$). This result is consistent with greater familiarity for recognized compared to unrecognized stimuli during RH-based decisions.

Non-adherent Trials

The early time window ANOVA for non-adherent trials did not result in any significant main effects of recognizability between two regions within a pair. The three-way interaction between condition, recognizability, and posteriority reported above was also not significant ($T_{W3}/c(1.0,34.0)=0.07, p=.79$), though it should be re-emphasized that this analysis was only run
on data from the FH conditions. A direct comparison of non-adherent ERPs to accordant ERPs for RH trials was not possible due to the low trial counts previously mentioned.

**500 - 800 ms**

*Adherent Trials*

The ANOVA for the time window corresponding to recollection (500 – 800 ms) revealed a significant three-way interaction between condition, recognizability, and posteriority \( (T_{WJt}\text{c}(2.0,41.8)=6.61, p=.005) \), such that there was only a two-way recognizability by posteriority interaction present for the FH large difference (> 400ms) condition \( (T_{WJt}\text{c}(1.0,47.0)=20.80, p<.001) \). Within the FH large difference trials, only the posterior ROIs yielded a significant effect of recognizability, such that ERPs in response to faster (more fluently) recognized regions within a pair were significantly more positive than ERPs in response to slower recognized regions \( (T_{WJt}\text{c}(1.0,47.0)=14.28, p<.001) \). This result is consistent with greater recollection for faster recognized regions compared to slower recognized regions during FH trials with large recognition latency differences.

*Non-adherent Trials*

The late time window ANOVA for non-adherent trials did not result in any significant main effects of recognizability between two regions within a pair. The three-way interaction between condition, recognizability, and posteriority reported above was also not significant \( (T_{WJt}\text{c}(1.0,34.0)=0.00, p=.99) \). Within the FH large difference (> 400ms) condition, the posterior ROIs did not yield ERPs significantly different for faster and slower recognized regions \( (T_{WJt}\text{c}(1.0,34.0)=0.64, p=.43) \). In contrast to adherent trials, there is no evidence of greater recollection for faster recognized regions compared to slower recognized regions during trials where participants’ decisions did not adhere to the FH.
300 - 800 ms

To examine possible differences between the early and late time windows, an ANOVA was conducted including the early and late time windows as an additional independent variable. Because the FH < 400 condition yielded no significant results in the above analyses, this condition was eliminated from the dataset for this analysis. Only adherent trials were analyzed.

There was a marginally significant four-way interaction between condition, time window, recognizability, and posteriority ($T_{WJt}/c(1.0,47.0)=3.83, p=.056$). If significant, this four-way interaction would potentially doubly dissociate familiarity at the early time window during RH trials at anterior electrode sites, and recollection at the late time window during FH trials (>400 ms) at posterior electrode sites. This clean dissociation seems held back by the finding that faster recognized regions in the FH (>400 ms) trials are significantly more positive than slower recognized regions at posterior sites during the early time window ($T_{WJt}/c(1.0,47.0)=6.59, p=.014$, see Figure 5 upper-right) as well as the late time window (Figure 5 lower-right), where they are expected. In other words, the fluency-related recollection effects may have started earlier in this experiment than is typical for recollection effects to start in recognition memory experiments.

Discussion

The present study implemented a city-size comparison task to investigate the impact of recognition memory on heuristic decision-making. Our method was to isolate recognition- and fluency-heuristic-based judgments and parse out the contributions of familiarity and recollection subserving both heuristics using ERPs. We predicted, based on the literature, that RH-based decisions would be associated primarily with a familiarity difference between recognized and unrecognized regions, as indexed by more positive FN400 effects for recognized regions.
Because a recognized and an unrecognized region should be strongly dissociable based on familiarity alone, we predicted that further cue knowledge would not be retrieved, and recollection would be indistinguishable for recognized and unrecognized regions, as indexed by parietal old/new effects. This prediction would support the noncompensatory claim of the RH, such that decisions would be based solely on an early familiarity signal without consideration of further recollection. The ERP results supported this prediction, as FN400 effects at anterior sites were significantly more positive for recognized regions than unrecognized regions. Parietal old/new effects at posterior sites, however, did not significantly differ between recognized and unrecognized regions. These findings are consistent with the framing of the RH as a fast, parsimonious, and noncompensatory decision mechanism. As suggested by Goldstein and Gigerenzer (2002), recognition, or more specifically familiarity, appears to serve as an initial screening step that can be used to differentiate two items. If two items are deemed reliably dissociable based solely on early familiarities, it seems logical that search for further cue knowledge (via recollection) can be abandoned and a decision made based solely on familiarity.

Goldstein and Gigerenzer claimed that further cue knowledge would only be searched for if both objects in a decision frame were recognized; a condition that would permit use of the FH but not the RH.

For inferences in Experiment 1 where both regions were recognized, and thus the FH could be applied, and there was a small retrieval fluency difference between regions, we predicted indistinguishable FN400 and parietal old/new effects. Based on their similar retrieval fluencies, these regions should be comparable in terms of how familiar they are and any recollected cue knowledge they elicit. The FH is not typically an advantageous decision-making strategy in this situation due to the similar fluency values, and previous studies have found that
people adhere less frequently to the FH in these cases (Hertwig et al., 2008; Schooler & Hertwig, 2005). Participants should largely resort to guessing in these situations, because there should be little to no information in memory that can provide a reliable decision cue. It is possible, if we assume two regions in this condition elicit similar amounts of recollection, that knowledge retrieved about one region provides a cue highly relevant to population (i.e., this region has an NBA team), whereas knowledge retrieved about another region is less relevant to population (i.e., my cousin lives in this region). In this case, participants could still rely on recollection to make a decision, but this would not be reflected in parietal old/new effects because both regions would elicit recollection of similar strengths.

For inferences in Experiment 1 with a large recognition latency difference between two recognized regions, our predictions were less concrete. The literature suggests a role for familiarity in these decisions, insofar as more fluent stimuli are more familiar. Hertwig et al. (2008) suggest that the FH can work through a conscious assessment of two retrieval fluencies, and imply that retrieval fluency is a reflection of relative familiarity for two objects. However, it seems likely that in situations where both regions are recognized, the corresponding familiarity signals would be more similar than when only one region is recognized, and familiarity alone might not be enough to dissociate faster recognized regions from slower recognized regions. The ERP results here appear to support this argument; there was no significant FN400 difference between faster recognized regions and slower recognized regions at anterior sites. This finding implies similar levels of familiarity for regions considered in FH-based decisions, even when there is a large retrieval fluency difference between the two regions. This result provides evidence against the notion that FH decisions are based on a familiarity comparison. Furthermore, the ERP results showed parietal old/new effects for faster recognized regions that
were significantly more positive than slower recognized regions, consistent with stronger recollection occurring for the faster recognized city. What type of information might be being recollected, and whether or not this information is being evaluated in supposedly FH-based decisions is difficult to assess. The most logical explanation is that participants are recalling factual details related to the given regions, and presumably could be using this information to guide their decisions, especially if there is no reliably exploitable distinction between the familiarity of two regions.

Taken together, our results corroborate existing perspectives of the RH and provide unique evidence for its assertion as a noncompensatory mechanism. Previous theoretical accounts have claimed that recognition is the lone cue considered in RH-based decisions, but fail to dissociate the relevant underlying memory processes. Rosburg et al.’s (2011) ERP results focused on a role for familiarity, yet also found significant parietal old/new effect differences (presumably indexing dissociable levels of recollection) between regions that were recognized and unrecognized. However, their methodological approach may have muddled RH trials with FH trials because recognized and unrecognized regions were determined a priori and were assumed to be universal for all subjects. The current study employed a subject-by-subject recognition test, separate from the population decision-making task, in order to purely dissociate RH and FH trials.

**Experiment 2**

The results of Experiment 1 suggest RH-based decisions are based solely on the familiarity component of recognition memory, and that recollection plays no discernable role on a majority of trials. There is more uncertainty surrounding the mechanisms at work behind the FH. Our ERP results suggest a distinction based on recollection differences between two regions,
with no contributions from familiarity. If true, and participants were actually utilizing recollected cue knowledge when making population decisions, this would indicate that perhaps the fluency cue is being ignored in favor of alternate knowledge regarding the regions. The apparent recollection-advantage for regions identified more quickly in the recognition test raises new questions, and in order to better assess how recollection might ultimately influence decisions, it might be best to revisit the core principles of the FH.

An initial important question to consider is how recollection might influence the speed at which an item is retrieved from memory - its retrieval fluency. One could make an argument for a longer retrieval fluency indexing either less or more recollection. Considering the first situation, longer retrieval fluency could correspond to a lengthier, more effortful search through memory when cue knowledge is sparse. Behaviorally, the relative unavailability of information regarding a certain item would lead to extended search and longer reaction times, which serve as a proxy for retrieval fluency. Alternatively, longer reaction times might reflect a higher volume of recollection. Lengthier retrieval fluencies would not index effortful search through memory, but rather an influx of readily available information, perhaps triggered through associations in memory. For example, Richardson-Klavehn (2010) has recently made an argument for a form of automatic recollection. He reviews certain memory experiments where more deeply encoded stimuli are associated with slower reaction times than shallowly encoded stimuli. In these situations, it is thought that more information regarding a well-known stimulus is automatically recollected upon retrieval, and therefore leads to longer retrieval times. However, Richardson-Klavehn goes on to review alternate experiments that demonstrate exactly the opposite - shorter reaction times associated with more deeply encoded stimuli. The key component that appears to separate these contrasting results is the task instructions. Tasks that instruct subjects to focus on
recollection itself, and either explicitly or implicitly place value on additional recollection, are associated with longer retrieval times for more deeply encoded stimuli. Tasks where retrieval is incidental and the instructions value some other process, more deeply encoded stimuli are associated with shorter retrieval times.

It is hard to say concretely how the instructions given in our recognition test from Experiment 1 might modulate reaction times (retrieval fluencies) for well-known compared to little-known regions. Subjects are simply asked to identify which regions they have heard of before, or are familiar with. These instructions do not place value on additional recollection, but they do ask subjects to directly access their memory. Our recognition test instructions seem more consistent with producing recollection as an incidental process of region identification, suggesting shorter retrieval fluencies for more well-known regions. In support of this interpretation, Hertwig et al. (2008) conducted an experiment that showed that reaction times in a recognition test were shorter for more populous regions, perhaps indicating that more well-known regions are associated with faster retrieval fluencies. Whether these recognition judgments were made on the basis of familiarity, recollection, or a combination of the two remains unknown.

Though it is logical to assume that regions identified more quickly have greater familiarity signals than those identified less quickly, our ERP results from Experiment 1 did not identify a reliable FN400 difference between these types of regions. Therefore, it is possible that faster retrieval fluencies are more reflective of stronger recollection for a given region. Shorter retrieval fluencies could index more instantaneous recollection of knowledge that dominates a familiarity distinction in situations where both stimuli are recognized. In an FH study conducted by Volz et al. (2010), they asked participants what strategies they used after taking part in a city-
size comparison task. For trials where both regions were recognized, participants said they relied on criterion knowledge (i.e., knowledge of the exact population, or population rank) when available, yet such knowledge was only available for a very small fraction of trials. When a decision could not be reached based on criterion knowledge, 17 of the 18 participants reported choosing the city that felt larger or more familiar, or that they had made an informed guess. These findings were interpreted as supporting a role for familiarity in guiding FH-based decisions. Importantly, however, it should be noted that Volz et al. did not mention whether their questionnaire gave participants the option of choosing the city that was associated with greater recollection or cue knowledge (other than exact population information, i.e., criterion knowledge).

However, if familiarity was indeed guiding FH-based decisions, we would have expected to see reliably divergent FN400 familiarity signals between the two recognized regions in Experiment 1. Because our ERP results suggest that two recognized regions are dissociated solely on the basis of recollection, it seems more likely that some sort of knowledge cues are being retrieved for each recognized stimulus, and subjects are basing their population decisions on which stimulus elicited greater or more useful recollection. In this way, retrieval fluency might be negatively correlated and confounded with recollection, making it difficult to determine which cue participants might be utilizing when making population decisions.

Experiment 2 aims to answer some of the questions surrounding retrieval fluency, familiarity, and recollection by collecting additional information about each region presented in the task. Because parietal old/new effect differences were not necessarily predicted a-priori in Experiment 1, Experiment 2 also seeks to validate these purported recollection differences behaviorally. Experiment 2 differed from Experiment 1 only in the recognition test phase. In
addition to providing timed “yes” or “no” recognition responses to each region in the task, a modified remember-know procedure was included. In the literature, the remember-know procedure has been commonly used to differentiate items that are familiar (“known”) from those that are consciously recollected (“remembered”). The procedure typically involves a study phase and test phase where participants are asked to make remember/know judgments pertaining to items in the previous study phase. Our implementation of this design instead focuses on pre-experimental memory, where participants make analogous remember/familiar judgments about cities or countries they had heard of outside the context of the experiment. Other researchers have used similarly modified versions of the remember-know procedure (Bird, Davies, Ward, & Burgess, 2011; Trinkler, King, Doeller, & Rugg, 2009) and typically found universally enhanced familiarity and recollection for pre-experimentally known items. For our version of the task, participants identified regions as “remembered”, “familiar”, or “unknown”. Additionally, for regions identified as “remembered”, participants provided information about how many specific contextual details they could recollect about that given region. Confidence judgments were obtained for all regions identified as “familiar” or “unknown”.

Regarding the relationship between retrieval fluency and familiarity/recollection, we predict that reaction times for regions identified as “familiar” will be longer than regions identified as “remembered”. Presumably, regions identified as ‘familiar’ are less well-known and not associated with any recollectable cue knowledge, so an effortful search through memory should persist longer in these cases in an attempt to recollect unavailable knowledge. Regions identified as “remembered” should be associated with faster reaction times as a result of readily accessible and recollectable context information. Following this same logic, we predict that regions where participants are able to identify higher quantities of distinct remembered details
will have faster reaction times than regions where participants identify lower quantities of distinct remembered details. Ideally, we would expect a continuum of reaction times, with the fastest recognized regions associated with the largest amounts of recollectable cue knowledge, and the slowest recognized regions associated with solely “familiar” judgments. This pattern of results would bolster our ERP findings from Experiment 1, such that faster recognized regions would be validated as more well-known, and provide direct evidence of the availability of a recollection-based cue that participants could be using as the basis for their population decisions in addition to the already-established fluency cue.

Collection of additional information in the recognition task will allow for finer-grain comparison of trials in the population decision-making task. In trials where the RH is applicable (i.e. one region is unknown and the other is “familiar” or “remembered”), we can predict that adherence rates and $r$-parameters (indicating sole reliance on recognition to making decisions) from the $r$-s-model should be similar in trials where the recognized region is simply “familiar” and those where the recognized region is “remembered”. If the RH is truly based on familiarity alone, as our results from Experiment 1 would suggest, we should not expect participants to behave any differently between {“unknown” vs. “familiar”} trials and {“unknown” vs. “remembered”} trials. Logically, any region that is remembered is also associated with some positive amount of familiarity that should be greater than the familiarity elicited from an unknown city, and RH decisions should be made based solely on the comparison of the two familiarity signals. In other words, recollection of additional information should not play a role in RH-based decisions.

Alternatively, based on our Experiment 1 results, we would expect recollection to play a larger role in situations where the FH is applicable (i.e. both regions are recognized). In trials
where there is a decipherable recollection difference between the two regions (i.e. a “remembered” region vs. “familiar” region, or a “remembered” region with a large amount of recollected knowledge vs. a “remembered” region with a small amount of recollected knowledge), the region with a recollection advantage should be chosen a majority of the time. However, regions with a recollection advantage are also likely to be more fluently retrieved, and therefore the retrieval fluency difference between two regions should roughly correspond to the recollection difference between two regions. By examining different trial types where the FH is applicable (e.g. {“familiar” vs. “familiar”}, {“familiar” vs. “remembered”}, {“remembered” vs. “remembered”}), we can use information provided by adherence rates, the s-parameters from the r-s-model (indicating sole reliance on the fluency cue), and logistic regression to hopefully parse out the differential contributions of retrieval fluency and recollection to purported FH-based decisions.

**Method**

**Participants**

Thirty-four participants (11 female) ranging in age from 18 to 23 were recruited to partake in the study. All participants were undergraduate students receiving course credit from the University of Colorado.

**Materials and Procedure**

Each participant performed two computerized tasks similar to those in Experiment 1: a city/country recognition test first and a population inference task second. Unlike Experiment 1, task order was not counterbalanced. Results from Experiment 1 yielded no significant effects of task order, and we wanted to obtain the purest measures of pre-experimental memory as possible. For example, if the recognition test were placed after the population decision making task,
participants might identify some regions as “familiar” after viewing them in the decision making task, whereas in reality these regions were “unknown” before beginning the experiment. Prior to beginning the experiment, each participant completed an approximately 3-min practice session for both tasks. Stimuli were identical to those used in Experiment 1: U.S. city and country names displayed in the center of a computer monitor. The stimuli that appeared in the practice session did not appear in the actual task.

For the recognition test, participants again viewed the 100 most populous cities in the United States, and the 100 most populous countries in the world, as well as 10 fictional cities and 10 fictional countries intended to increase the honesty of responses. Order of city and country blocks was counterbalanced. Each trial began with a 2 s fixation cross (+), followed by a single randomly selected region name on the center of the screen. On the first screen, participants were instructed to indicate with a “yes”/“no” button press whether or not they recognized each region from prior to the experiment, just as they were in Experiment 1. Reaction times were recorded for this first response and interpreted as the recognition latency for that given region. Key assignments appeared at the bottom of the screen, and order of key assignments was counterbalanced. After the first response was made, the stimuli remained on the screen, but the key assignments at the bottom of the screen updated to a three-choice set (“Remember”, “Familiar”, “Unknown”). During this second screen, participants were instructed to identify whether they could “remember” each region, described as recall of any type(s) of specific details about that region from prior to the experiment; the region was “familiar”, described as knowing they have heard of that region prior to the experiment, but being unable to recall any specific details; or “unknown”, described as never having heard of that region before. Stimuli remained on the screen until a response was made, and accuracy was emphasized over speed. If
participants identified a region as “remembered”, they were immediately prompted with the question “How many details can you recall about [region X]?” on the center of the screen. Response options appeared below, with four choices ranging from 1 to 4+ (4 or more), and their counterbalanced key assignments beneath them. Responses were untimed, and upon making a choice the trial ended and the next trial began with a 2s fixation cross. If participants instead identified a region as “familiar” or “unknown”, they were immediately prompted with the question “How confident are you that [region X] is [familiar/unknown] to you?” on the center of the screen. Response options appeared below, as a set of four confidence choices (“Guess”, “Minimally”, “Somewhat”, “Very”), with their designated counterbalanced key assignments below. Confidence judgments for “familiar”/“unknown” responses were primarily included in an attempt to equalize participant effort across all trial types. Upon making a confidence judgment, the trial ended and the next trial began with a 2s fixation cross.

The population inference task for Experiment 2 was nearly identical to that described in Experiment 1 (See Figure 2). Participants viewed two regions sequentially, and were then asked to choose which region they thought had a larger population. The only deviation from Experiment 1 was that participants viewed each region for 1500 ms instead of the previous 2000 ms, because EEG was not recorded and longer stimulus durations were no longer necessary.

**Results**

Participants recognized on average 79 out of the 100 most populous U.S. cities, and 86 out of the 100 most populous countries. This resulted in an average of 272 FH-applicable trials, 104 RH-applicable trials, and 17 guessing trials (neither region recognized) per participant. Additionally, participants on average only claimed to recognize 1.7 of the 20 fictional cities and countries included in the recognition test (no difference for cities versus countries), so responses were honest. The recognition validity \( (M = .76, \ SD = .05) \) and fluency validity \( (M = .59, \ SD = \)
.05) were within range of previously reported findings (Hertwig et al., 2008; Hilbig et al., 2011) including our own Experiment 1 findings ($M = .76, M = .59$, respectively), and both were significantly greater than chance ($t(33) = 28.7, p < .0001; t(33) = 10.4, p < .0001$; respectively) indicating that recognition was an ecologically rational cue during RH trials, and fluency was an ecologically rational cue during FH trials. Similar to Experiment 1 analysis, adherence rates can be calculated separately for RH trials and FH trials in order to assess how frequently participants’ inferences were in line with each heuristic’s predictions. The adherence rate for the RH ($M = .85, SD = .10$) was significantly above chance ($t(33) = 19.6, p < .0001$), indicating that participants’ choices were in line with the RH’s prediction on a majority of trials. Additionally, the adherence rate for the FH ($M = .62, SD = .05$) was significantly above chance ($t(33) = 14.1, p < .0001$), indicating that participants’ choices were in line with the FH’s prediction on a majority of trials.

*Multinomial Processing Tree analysis*

Thirty-four participants resulted in an aggregate of 13,383 inference trials (see Appendix B). Similar to Experiment 1, considering the large number of trials and high statistical power for a goodness of fit test, the model fit the data well ($G^2(1) = 3.14, p = .08$). Parameter estimates are displayed in Table 2 below.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Psychological Meaning</th>
<th>Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>recognition validity</td>
<td>.76 (.01)</td>
</tr>
<tr>
<td>$b1$</td>
<td>knowledge validity, fluency-homogenous FH cases</td>
<td>.63 (.01)</td>
</tr>
<tr>
<td>$b2$</td>
<td>knowledge validity, fluency-heterogeneous FH cases</td>
<td>.67 (.01)</td>
</tr>
<tr>
<td>$b3$</td>
<td>knowledge validity, RH cases</td>
<td>.66*</td>
</tr>
<tr>
<td>$c$</td>
<td>fluency validity</td>
<td>.61 (.01)</td>
</tr>
<tr>
<td>$g$</td>
<td>correct guessing (neither object is recognized)</td>
<td>.51 (.02)</td>
</tr>
<tr>
<td>$p$</td>
<td>proportion of fluency-homogenous FH cases</td>
<td>.13 (.00)</td>
</tr>
<tr>
<td>$r$</td>
<td>RH-use (considering the recognition cue in isolation)</td>
<td>.63 (.01)</td>
</tr>
</tbody>
</table>
Table 2. Parameters of the r-s-model, psychological meaning of the parameters, and parameter estimates with standard errors of each estimate, based on data from all participants.

| s    | FH-use (considering retrieval fluency in isolation) | .21 (.01) |

This number is derived analytically from $b_3 = p \times b_1 + (1 - p) \times b_2$ and is thus reported without a standard error.

The model estimated a recognition validity identical to that reported in the observational statistics above ($M = .76$), as well as a nearly identical fluency validity ($M = .61$), thus corroborating the remaining estimates obtained from the r-s-model. Again, the two parameter estimates of greatest importance are the probability of RH-use based on recognition alone ($r$ parameter) and the probability of FH-use based on retrieval fluency alone ($s$ parameter).

According to the r-s-model, participants relied on the recognition cue in isolation when one region was recognized and the other was not recognized on 63% of the trials. This estimate is lower than the mean adherence rate reported above ($M = .85$), though still significantly greater than chance ($\Delta G^2 = 67, p < .0001$, when fixing $r = .50$), indicating that participants relied on the RH on a substantial portion of trials.

Most importantly for the given research question, the r-s-model estimated that participants relied on retrieval fluency in isolation when both regions were recognized on only 21% of the trials ($s = .21$). This estimate replicates Hilbig et al.’s (2011) findings, indicating that participants only used the FH on approximately one fifth of the trials in which it could have been applied. This estimate, which unconfounds the contributions of retrieval fluency and further knowledge or other information, is significantly lower than the mean FH adherence rate reported above ($\Delta G^2 = 1713, p < .0001$, when fixing $s = .62$). So, across all participants, use of the FH seemed to be quite sparse.

Previous literature on heuristic decision making has emphasized the need to assess data on the individual level in addition to the aggregate level, particularly because different
individuals may indeed differ in their use of conditional heuristics (Goldstein & Gigerenzer, 2002; Hertwig et al., 2008, Hilbig et al., 2008, among others). To test if the reported findings above hold on an individual level, we applied the r-s-model to each participant’s data to obtain individual parameter estimates. Results indicated that the r-s-model fit 32 out of the 34 participants’ data well ($G^2 < 4, p > .05$), with the remaining two participants obtaining a reasonable fit ($G^2 < 10, p > .002$). Figure 7 shows individual estimates of the probability of RH-use ($r$ parameter) as compared to the corresponding RH adherence rate for each participant.

![Figure 7](image)

**Figure 7.** Individual probabilities of using the Recognition Heuristic (RH) for each participant. Gray bars indicate the probability of RH-use implied by adherence rates, and red bars indicate the probability of RH-use implied by the $r$ parameter, with participants sorted by adherence rate.

As is evident in Figure 7, no participant had an $r$ parameter larger than their adherence rate, and thus the finding of reduced RH-use reported from the aggregate results appears to represent a pattern across all participants. A t-test comparing these values also indicated that on average, adherence rates were significantly higher than $r$-parameters ($t(66) = 4.85, p < .0001$),
Figure 8 show individual estimates of the probability of FH-use ($s$ parameter) as compared to the corresponding FH adherence rate for each participant.

![Graph](image)

**Figure 8.** Individual probabilities of using the Fluency Heuristic (FH) for each participant. Gray bars indicate the probability of FH-use implied by adherence rates, and red bars indicate the probability of FH-use implied by the $s$ parameter, with participants sorted by adherence rate.

Similarly for the FH, no participant had an $s$ parameter larger than their adherence rate, and furthermore a majority of participants exhibited adherence rates more than double their estimated $s$ parameter. Importantly, no participant relied on retrieval fluency alone in greater than 38% of the trials, suggesting that all participants relied on more than simply retrieval fluency to make choices on a majority of trials. Figure 8 illustrates that the finding of reduced true FH-use reported in the aggregate results is representative of a pattern across all participants, and not driven by a select group of individuals. A t-test confirmed $s$-parameters significantly lower than adherence rates on average ($t(66) = 23.46$, $p < .0001$).

**Remember-Know analysis**
In addition to recognition responses and retrieval fluencies, information pertaining to a participant’s perceived memory for each region was collected through the modified remember-know procedure described above. Regions could be identified as remembered, familiar, or unknown (for the first analysis we will focus only on regions identified as remembered or familiar). Furthermore, each region identified as remembered was associated with one through four or more specifically recalled details relevant to that particular region. This resulted in a total of five possible perceived memory categories for each region: familiar (F), remembered with 1 detail (R1), remembered with 2 details (R2), remembered with 3 details (R3), and remembered with 4 or more details (R4).

To assess the relationship between retrieval fluency and perceived memory, mean recognition latencies for the five memory conditions were as follows (see Figure 9): $M(F) = 1832$ ms, $M(R1) = 1570$ ms, $M(R2) = 1476$ ms, $M(R3) = 1349$ ms, $M(R4) = 1148$ ms.

![Figure 9](image_url) **Figure 9.** Boxplot of aggregate recognition latencies (RTs) prior to log transformation for all regions on the y-axis, grouped by the five memory categories of interest on the x-axis. Solid horizontal lines within boxes indicate the average recognition latency for each memory category, and responses above tails indicate responses in the upper 25\textsuperscript{th} percentile for each memory category.
A linear mixed model was fit to the data, using perceived memory as the categorical independent variable and recognition latency as the continuous dependent variable. Reaction times were log-transformed prior to analysis. The model resulted in a significant linear effect of memory strength on recognition latency ($r = -.701$), indicating that as perceived memory strength for a region incrementally increased, recognition latency decreased ($F(1, 32.86) = 80.47, p < .0001$). $F$-statistics and $p$-values were obtained using the Kenward-Rogers approximation (Kenward & Roger, 1997). This result demonstrates that retrieval fluency (i.e., recognition latency) is indeed confounded and negatively correlated with the amount of further knowledge accessible for a given region (i.e., perceived memory). It also demonstrates that regions identified more quickly are directly associated with greater knowledge or recollection, and corroborates the existence of a ‘recollection cue’ that participants could capitalize on when making population decisions.

To consider the usefulness of perceived memory and further knowledge as a potential decision cue during the inference task, we can examine trials where both regions were identified as remembered or familiar. Figure 10 shows inference task choices for trials where one region was identified as remembered and one region was identified as familiar (RvF trials) on the left, and inference task choices where both regions were identified as remembered (RvR trials) on the right, for each participant.
In trials where one region was identified as remembered and the other was identified as familiar (Fig. 10A), participants chose the remembered region as being more populous 77% of the time, significantly above chance level ($t(33) = 14.18$, $p < .0001$). It should be noted that this number is substantially larger than the overall FH adherence rate ($M = .62$), suggesting that subjective memory for a region is potentially a more useful decision cue than retrieval fluency (however, these two results cannot be directly compared because they consist of the same overlapping observations). In trials where both regions were identified as “remembered” (Fig. 10B), we can restrict analysis to pairs where one region was identified as relatively strongly remembered (3-4 or more details remembered) and the other region was relatively weakly remembered (1-2 details remembered). In these trials, participants chose the more strongly remembered region as being more populous 70% of the time, significantly above chance level ($t(30) = 9.72$, $p < .0001$). This result suggests that even when specific details can be recalled for both regions within a pair, the

**Figure 10.** (A) Inference choices for remembered vs. familiar fluency heuristic trials. Participants are sorted on the x-axis, with the percentage of trials they chose the remembered region as being more populous than the familiar region on the y-axis, with the red line indicating 50%. (B) Inference choices for remembered vs. remembered fluency heuristic trials. Participants are sorted on the x-axis, with the percentage of trials they chose the more strongly remembered region (3-4 memory details) as more populous than the less strongly remembered region (1-2 memory details), with the red line indicating 50%. Subjects with less than 5 trials in...
region associated with greater recollection is typically chosen as being more populous, and chosen more frequently than the region that was simply retrieved more speedily.

In order to parse out the relative contributions of recognition latency and perceived memory strength to population decisions, we adopted a mixed model approach based on the aggregate data. A memory difference variable was created for each pair of regions in the inference task by subtracting the value of region 2 from region 1 (familiar = 0, R1 = 1, R2 = 2, R3 = 3, R4 = 4), forming a 5-level categorical variable. Retrieval fluency difference between both regions within a pair was also computed by subtracting the recognition latency of region 2 from region 1. An initial model incorporated all trials where either region was identified as familiar or remembered (4,790 trials), and used log-transformed recognition latency difference and perceived memory difference to predict participants’ decisions. This model yielded a strong simple effect of perceived memory difference ($b = -.40, SE = .03, p < .0001$), as well as a strong simple effect of recognition latency difference ($b = 2.28, SE = .31, p < .0001$). However, this model also incorporates comparisons where both regions received the same perceived memory response (e.g. familiar vs. familiar), and therefore perceived memory difference is an uninformative predictor of choice. Focusing solely on trials where one region was identified as familiar, and another was identified as remembered (R1-R4); we fit the same mixed model. Again, there was a strong simple effect of perceived memory difference ($b = -.42, SE = .04, p < .0001$), and a slightly attenuated, though still significant simple effect of recognition latency difference ($b = 1.92, SE = .62, p = .002$). These results suggest that both retrieval fluency and

---

3 Recent research surrounding linear mixed models has warned against reporting degrees of freedom and $F$-statistics, due to the fact that the pivotal quantities for these tests do not have $t$ or $F$-distributions (e.g., Baayen, Davidson, & Bates, 2008). As alternative, it is encouraged to report parameter estimates, standard errors, and $p$-values when sample size is large enough to justify, as it is in our sample.
further knowledge (via perceived memory) play differential roles when participants are making population decisions.

Returning to the multinomial processing model (r-s-model), we can now look at different trial types based on perceived memory and observe their effects on RH and FH use. For RH cases, we can examine “remembered” vs. “unknown” (RvU) trials and “familiar” vs. “unknown” (FvU) trials. To do this, we can extract the RH tree within the r-s-model, and compare only RvU and FvU observations as separate trees within a new model. This procedure is valid because trees within multinomial processing models are independent (Batchelder & Riefer, 1999), and we have established that the model holds for our data. When comparing these two trees, it is not possible to perform a goodness of fit test (the new model is saturated), but this is not a problem because we have shown that the full model holds with our data. However, we cannot definitively show that the full model holds for the separation of RvU and FvU trials, thus we cannot rule out the possibility that the full model operates differently in these cases, and therefore our comparison of output from these different trial types in the new model cannot be taken as indisputable.

The comparison of RvU and FvU trials within the new model output recognition validity estimates of .80 and .70, respectively. This finding suggests that, because the probability of the recognized region actually being more populous is higher in RvU cases, that the recognition cue is more valid in RvU cases. Furthermore, the new model output r-parameter estimates of .71 for RvU trials and .59 for FvU trials, indicating participants relied on mere recognition alone to make their decisions in 71% of RvU trials and 59% of FvU trials. By setting this model as a baseline model, and fixing the r-parameter to be a constant 71% for FvU trials (or alternatively setting the r-parameter to be a constant 59% for RvU trials) in a new model, we can statistically compare these two r-parameters. This model comparison indicates that the r-parameter for RvU
trials is significantly greater than the \( r \)-parameter for FvU trials \( (\Delta G^2 = 26.2, p < .0001; \text{ when fixing } r_{FvU} = .71) \), implying that participants were using the RH on a greater portion of RvU trials than FvU trials. A unique latent parameter this model estimates (that is also provided by the r-s-model) is knowledge validity, or the probability of retrieving valid knowledge as opposed to invalid knowledge. The knowledge validity for RvU trials was .65, and the knowledge validity for FvU trials was .59. A comparison of these values indicates that knowledge is a more valid cue on RvU trials compared to FvU trials \( (\Delta G^2 = 5.09, p = .024; \text{ when fixing } b_1 = .59) \). Taken together, this output may appear to provide evidence for a role of recollection in RH-based decisions, such that trials where knowledge was available were associated with greater RH use, but a potential contribution of familiarity to this difference should also be considered due to the continuous nature of familiarity. We will return to this consideration in the general discussion.

In addition to examining different RH trial types, there are three different FH trial types that can be examined: “remembered” vs. “familiar” (RvF), “remembered” vs. “remembered” (RvR), and “familiar” vs. “familiar” (FvF). Just as was done with the RH, we can extract the FH tree from the r-s-model and create a new model composed of three separate FH trees to accommodate the three separate trial types within one model. The knowledge validities of the three trial types were estimated to be: RvF, .72; RvR, .67; FvF, .60. These numbers indicate that knowledge, or recollection, was the most valid cue for RvF trials, followed by RvR and FvF trials. This finding is consistent with recollection being most useful for RvF trials, where recollection only occurs for one region within the pair. The \( s \)-parameter estimates were as follows: RvF, .276; RvR, .174; FvF, .156. These numbers indicate that retrieval fluency was used in isolation to make population decisions on 27.6% of RvF trials, 17.4% of RvR trials, and 15.6% of FvF trials. We can statistically compare these numbers in a model comparison fashion.
similar to what was done for RH trials. This comparison indicates that the retrieval fluency cue was used in isolation on a significantly larger portion of decisions for RvF trials than RvR trials ($\Delta G^2 = 29.1 \ p < .0001$; when fixing $s_{RvF} = .174$) and FvF trials ($\Delta G^2 = 39.8 \ p < .0001$; when fixing $s_{RvF} = .156$). True FH use for RvR trials and FvF trials did not significantly differ ($\Delta G^2 = 0.43 \ p = .51$; when fixing $s_{RvR} = .156$). These numbers indicate that people are actually using the FH more often on RvF trials, where we would predict further knowledge to also be most useful, than RvR and FvF trials. However, the r-s-model has a built-in retrieval fluency threshold of 100 ms, meaning that trials where two regions have retrieval fluencies less than 100 ms apart from each other are considered to be homogenous and are excluded from the input to the generated s-parameter. The model output reports the proportion of trials where two regions are indecipherable based on fluency (the $p$ parameter) for each condition: FvF, .131; RvF, .089; RvR, .159. This output indicates that the fluency cue was generally more available on a greater proportion of RvF trials than FvF trials ($\Delta G^2 = 49.85 \ p < .0001$; when fixing $p_{RvF} = .131$) and RvR trials ($\Delta G^2 = 123.68, \ p < .0001$; when fixing $p_{RvF} = .159$). In addition to availability of the fluency cue, the model also outputs estimates of fluency validity (proportion of trials the more quickly recognized region is actually more populous) for the three trial types: FvF, .562; RvF, .643; RvR, .593. A comparison of these numbers shows that fluency validity is greater for RvF trials than both RvR ($\Delta G^2 = 42.86, \ p < .0001$; when fixing $c_{RvR} = .643$) and FvF ($\Delta G^2 = 38.55, \ p < .0001$; when fixing $c_{FvF} = .643$) trials. This suggests that retrieval fluency, or recognition speed, is most useful as a decision cue in RvF trials compared to RvR and FvF trials. The potential implications of these findings will be discussed further in the discussion.

**General Discussion**
The main idea behind the recognition and fluency heuristics is that in certain situations, decision makers are able to capitalize on recognition (in the case of the RH), or speed of retrieval (in the case of the FH), and use this memory-based information as a cue when choosing amongst objects. However, extant research has made limited progress towards connecting these decision-making heuristics to the specific hypothesized memory processes. Recent years have seen these heuristics challenged from multiple angles (Bröder & Eichler, 2006; Hilbig 2010; Hilbig et al., 2011; and Newell & Fernandez, 2006, among others), and as such a better understanding of the fundamental underlying processes could make steps toward resolving the controversy surrounding the RH and FH as realistic and practical models of comparative judgment.

In two experiments we adopted a dual-process perspective of recognition memory aimed to uncover the different memory components at play in the RH and FH. Although previous studies have hinted at a role for familiarity in both heuristics (Hertwig et al., 2008; Hilbig et al., 2011; Rosburg et al., 2011; Volz et al., 2010), its complementary process, recollection, has been largely ignored. Findings from both experiments supported a role for familiarity in RH-based decisions, indicating that participants could capitalize on early familiarity differences between two regions to make population decisions. Conversely, both experiments supported a role for recollection in FH-based decisions, indicating that in situations where there was not a reliable familiarity difference between regions, further memory search via recollection could provide knowledge cues that participants could capitalize on to make population decisions. We will first review findings concerning each heuristic, and then discuss the theoretical implications these findings have on the understanding of the recognition and fluency heuristics.

**The Recognition Heuristic**
Results from Experiment 1 suggested that during RH-based decisions, participants could rely on familiarity (as indexed by the FN400 ERP) to make their population decisions. The lack of differential recollection between regions (as indexed by parietal ERP old/new effects) during RH-based decisions provided evidence for the noncompensatory nature of the RH, insofar as recollection did not differ between regions and therefore could not be used as a reliable cue when making decisions. Results from the multinomial processing model corroborated this finding, indicating that participants relied on recognition in isolation, ignoring any further knowledge, to make decisions on approximately 76% of the trials in Experiment 1.

Experiment 2 mimicked Experiment 1, with the addition of a remember-know procedure in the recognition test in order to obtain more extensive behavioral measures of memory for each region in the study. The addition of behavioral measures allowed us to demonstrate that in the absence of recollectable cue knowledge, RH-based decisions could be made based solely on familiarity (RH use during FvU trials = 59%). This finding provided behavioral evidence in support of the interpretation of significant FN400 familiarity effects observed in Experiment 1.

However, our findings regarding the two possible RH trial types (RvU and FvU) did not match our predictions. We expected RH use (as measured by the $r$-parameter in the r-s-model) to not significantly differ for RvU vs. FvU trials. Our theoretical stance thus far has been that familiarity processes are utilized in RH decisions, and that recollection plays no discernible role. By this viewpoint, we would predict similar $r$-parameters (indicating sole reliance on familiarity to make decisions, ignoring further knowledge) for the two trial types. A “familiar” region and a “remembered” region are both associated with a positive familiarity signal, which should be utilized to make population decisions, ignoring any further recollection in a noncompensatory fashion. However, what we find is an $r$-parameter of .59 for FvU trials, significantly lower than
the $r$-parameter of .71 for RvU trials, implying greater RH use during RvU trials. This result could be taken to suggest that a sizable familiarity difference between two regions within a pair was more frequently available during RvU trials, when the recognized region was “remembered”, compared to FvU trials, when the recognized region was simply “familiar”. By the nature of the $r$-$s$-model, these estimates still provide evidence that participants are relying solely on recognition (familiarity), and not further knowledge (recollection), on a majority of trials in both situations, so the RH still accounts for a substantial portion of participants’ decisions.

A more elaborate assessment of familiarity during RvU and FvU trials can help us hypothesize why the two trial types resulted in differential amounts of RH use. Specifically, “remembered” regions could have greater familiarity than “familiar” regions, because they should be associated with more frequent exposure in one’s environment, and familiarity is known to be a continuous process (Woodruff et al., 2006; Yonelinas, Otten, Shaw, & Rugg, 2005). Therefore, there should be a larger discrepancy between the familiarity of {“remembered” vs. “unknown”} regions and {“familiar” vs. “unknown”} regions. This means the familiarity cue is more robust in the former cases (the two familiarities are more disparate), and can be utilized on a greater proportion of trials. Alternatively, the familiarity difference between two regions in {“familiar” vs. “unknown”} cases should be more homogenous, rendering the familiarity cue less useful. Thus, the finding of greater RH use for RvU trials compared to FvU trials could be explained by the fact that the familiarity cue is more useful and more robust during RvU trials compared to FvU trials.

The Fluency Heuristic
During FH-based decisions, there were no reliable ERP FN400 familiarity differences between more quickly and slowly recognized regions, though more quickly recognized regions were associated with greater ERP parietal old/new recollection effects than more slowly recognized regions. This finding implied that in the absence of reliably differential familiarity between two regions, later recollection processes provided information that could be capitalized on when making population decisions. Although the FH by definition assumes that decision makers are relying on consciously assessable retrieval fluencies to make choices, our Experiment 1 findings suggest the presence of alternate knowledge cues that could presumably be utilized to make population decisions. Results from the multinomial processing model indicated that participants relied on retrieval fluency in isolation to make their decisions on only 16% of trials, lending further credence to the idea that participants were instead capitalizing on recollected knowledge to make their decisions.

However, this study highlighted an apparent confound between retrieval fluency and recollection: more fluently retrieved regions were also associated with greater parietal old/new effects thought to index greater recollection. In Experiment 2, it was necessary to test whether these more speedily recognized regions were actually associated with greater amounts of recollected knowledge. We found, as we predicted, that regions identified as simply “familiar” were associated with the slowest retrieval fluencies, and that retrieval fluencies linearly decreased as participants identified greater amounts of recollectable cue knowledge for “remembered” regions. This finding suggests that recognition of regions that are simply “familiar” induces a more time-consuming and effortful search through memory in an attempt to recollect unavailable knowledge, whereas recognition of regions associated with large amounts of recollectable knowledge is more instantaneous, perhaps due to an abundance of readily
accessible information. Because this result shows that faster retrieved regions are associated with more recollectable knowledge, it provides direct behavioral evidence in support of our ERP FH results from Experiment 1, demonstrating the availability of a recollection-based distinction between more quickly and slowly retrieved regions that participants could capitalize on when making population decisions.

FH trials from Experiment 2 can be further examined with the s-parameters (indicating sole reliance on fluency/recognition speed to make population decisions). The three possible trial types for the FH are: {“remembered” vs. “remembered”; RvR}, {“remembered” vs. “familiar”; RvF}, and {“familiar” vs. “familiar”; FvF}. Based on our theoretical assertions, recollection processes are being utilized during FH decisions, and therefore the recollection cue should be most useful on trials with larger recollection differences between pairs of regions. If both decision cues (recollection and retrieval fluency) are viable options, but the recollection cue is indeed superior to the fluency cue, we would only expect retrieval fluency to be utilized in situations where the recollection cue provides little information. These situations should be FvF trials, and to a lesser extent RvR trials. Theoretically, there is no recollection available for either region in FvF trials, so on these trials participants should either rely on the fluency cue or resort to guessing. On RvR trials, there might only be a reliable recollection difference between regions when one region is more strongly recollected (3-4 specific details are remembered) and the other is more weakly recollected (1-2 specific details are remembered). Here, participants can rely on the recollection cue if it is valid, and if not can fall back on the fluency cue to make a decision or again resort to guessing. RvF trials should always provide a reliable recollection cue, because one region is “remembered” with some level of recollection occurring, and the other region is “familiar” with presumably no recollection occurring. On these trials, we would expect very little
reliance on the fluency cue, because recollection should dominate the fluency cue on all trials. The s-parameters we found for each trial type are as follows: FvF, 16%; RvF, 28%; RvR, 17%. For FvF and RvR trials, the r-s-model indicates participants are utilizing the fluency cue very rarely, at only a 16-17% rate. Alternatively, during RvF trials, the model indicates participants are using the fluency cue slightly more than one-quarter of the time at 28%, and significantly more often than RvR and FvF trials. These results are contradictory to our predictions, which posited RvF trials should have the lowest s-parameters, followed by RvR trials and FvF trials with slightly higher s-parameters.

A possible explanation for these results lies in the previously demonstrated confounding between retrieval fluency and recollection. Put simply, trials with larger recollection differences should be the same trials that have larger retrieval fluency differences. Therefore, as the recollection cue becomes more valid, the fluency cue becomes more valid as well. The r-s-model also reports the proportion of trials where fluency is considered indecipherable between regions (< 100 ms), which are excluded from the input to the s-parameter. The model output reports the proportion of trials where this is the case for each condition: FvF, 13.1%, RvF, 8.9%, RvR, 15.9%. These numbers indicate that the fluency cue fails to distinguish between two regions on a significantly larger portion of FvF and RvR trials compared to RvF trials. This finding implies that the fluency cue is simply more available for RvF trials, and therefore more likely to be utilized. Beyond simple availability, the r-s-model also outputs the fluency validity for each trial type. Recapping, fluency validity refers to the proportion of trials where the more fluently retrieved region is actually more populous, regardless of which region participants actually choose. RvF trials also have a significantly higher fluency validity (64%) compared to FvF (56%) and RvR (59%) trials. Taken together, these two measures indicate that the fluency cue is
not only most available during RvF trials, but it is also most valid. This could help explain why we find a greater reliance on the fluency cue in situations where recollection should be most helpful (RvF trials) - because the fluency cue is also most helpful. Ultimately however, reliance on fluency in isolation occurred on 28% of RvF trials, still substantially below the majority, leaving open the possibility of reliance on recollection for the remaining 72% of trials.

To investigate this possibility, we examined trials where there was a reliable recollection difference between the two regions in a pair (i.e. a “familiar” region vs. a “remembered” region, or a region “remembered” with 3-4 or more details vs. a region “remembered” with only 1-2 details). If participants were utilizing the recollection cue, they should be choosing the region associated with stronger recollection as being larger. In cases where one region was “remembered” and the other “familiar”, participants chose the “remembered” region on 77% of trials, and in cases where one region was associated with stronger recollection and the other with weaker recollection, participants chose the more strongly recollected region on 70% of the trials. Although a statistical comparison is not possible due to overlapping observations, it should be noted that these numbers are substantially larger than the overall FH adherence rate (62%), a biased measure to begin with, suggesting that participants followed the recollection cue more frequently than the fluency cue.

Output from the multinomial processing model as well as choice data from trials with reliable recollection differences currently point to the recollection cue as being superior to the fluency cue. To address this question from another angle, we created a mixed model with the recollection cue and fluency cue as separate independent variables to predict participants’ population decisions. Across all trials in the aggregate data, both cues accounted for unique and substantial portions of the variance. However, after restricting analyses to trials where the two
regions within a pair differed on both predictors (i.e. removing trials where both regions have the same retrieval fluency, and trials where both regions have the same memory responses (e.g. “familiar” vs. “familiar”)), the simple effect of the retrieval fluency cue was attenuated and the simple effect of the recollection cue was strengthened. This result suggests a more favorable reliance on recollection when making decisions in situations where both predictors (fluency and recollection) are useful. Taken together, however, the present analyses suggest separate and meaningful roles for recollection and retrieval fluency in decisions where the FH is applicable, although current evidence suggests that recollection is the superior cue.

**Theoretical Implications for the RH and FH**

When our data from both Experiment 1 and Experiment 2 were fit to Hilbig et al.’s (2011) r-s-model, it returned results that closely replicated their main finding: retrieval fluency in isolation was only used in approximately one-fifth of the trials where it could be applied. This outcome came in contrast to the model’s estimates of RH use, which posited that the RH accounted for a substantial portion of participants’ decisions. Such low estimates of FH use obtained by Hilbig et al. caused them to question the plausibility of the FH as valid model of comparative judgment. If the RH and FH rely on similar underlying processes, why is the RH robustly outperforming the FH? Hilbig et al. pointed to a disconnect between the recognition memory literature and fast and frugal heuristics literature in their respective interpretations of “fluency” and “recognition” in an attempt to explain the starkly different performances of each heuristic in the r-s-model. We will first summarize these inconsistencies, and then discuss the theoretical implications of our findings for both heuristics.

Fluency has been an elusive and mischievous term strewn across several domains of psychology research. In the area of recognition memory, fluency is assumed to be experienced as
a heightened sense of familiarity, which, in turn, can be used for a recognition judgment in the absence of actual recall (Jacoby & Dallas, 1981). So, in terms of dual-process theories of memory, as Hilbig et al. (2011) point out, this path from fluency via subjective familiarity is typically considered the alternative route to recognition as opposed to conscious recollection (Jacoby, 1991). According to the recognition memory literature, Hilbig et al. argue, recognition and fluency are inherently intertwined via familiarity, and would therefore exert their potential influence on decisions in unison.

Alternatively, a different view is held within the fast and frugal heuristics program, which posits separate and distinct judgment strategies that come with their own special set of triggering conditions (Gigerenzer, 2004). From this viewpoint, recognition and fluency are treated separately, with each forming the basic cue for the RH and FH respectively. Furthermore, the FH is actually conditional upon recognition, such that fluency only exerts its influence on a decision if both items within a pair are recognized. However, Hilbig et al. (2011) point out that theories of recognition memory assert that recognition and fluency actually share what could be called a continuous familiarity variable as a common denominator (Westerman, Miller, & Lloyd, 2003; Whittlesea & Leboe, 2003).

The root of the disconnect between these two bodies of research, as Hilbig et al. (2011) seem to imply, is the interpretation of recognition as binary (yes or no) from the fast and frugal heuristics perspective, and more of a continuous process (via underlying familiarity) from the recognition memory perspective. If recognition is considered as more of a continuous, nonbinary cue (Erdfelder, Küpper, & Mattern, 2011), comparative judgments could be performed based on this familiarity-driven cue without the need for two separate heuristics (Shah & Oppenheimer, 2008). Hilbig et al. advocated collapsing the RH and FH into a single heuristic, because by
allowing both RH and FH decisions to be made based on a continuous familiarity cue, you eliminate the need for two distinct heuristics that rely on two separate cues. Although we believe that Hilbig et al. make several valid points, our data presents challenges to this perspective. Though the authors initially adopt a dual-process perspective of memory, noting the relationship between familiarity and fluency, no hypotheses are put forth regarding a potential role for recollection. Through their understanding, recognition is functionally synonymous with familiarity. By neglecting the potential impact of recollection, the authors might be overvaluing the role of familiarity. Indeed, claiming that all decisions, regardless of whether they fall under the RH or FH umbrella, are made on the basis of an underlying familiarity-driven signal is placing a heavy burden on familiarity, and assuming a high sensitivity to this signal. Is our perception of familiarity acute enough to accurately discriminate minor differences, or might later memory processes be stepping in to bail out familiarity?

Based on our data from the above experiments, we do not think familiarity can tell the whole story. However, much of Hilbig et al.’s (2011) account is in line with what we believe is occurring during RH-based decisions. Returning to our ERP results from Experiment 1, we found that familiarity signals could accurately discriminate recognized regions from unrecognized regions, with no contribution from later recollection processes. This is not unexpected; RH cases should inherently necessitate a rather large familiarity difference due simply to the fact that one region is completely unknown, whereas the other is associated with, at minimum, some marginal level of familiarity. These cases are cherry-picked, by the definition of the RH, to create sizeable familiarity differences. Because an early, automatic familiarity process can dissociate between two items, it is not necessary to pursue further recollection. We did not find significant parietal old/new effect differences for RH-based decisions, though there was a
slight trend in the direction of greater recollection for recognized compared to unrecognized regions. However, Rosburg et al. (2011) did report significant parietal old/new effects during RH-based decisions, albeit at an earlier time window than is typical and with a slightly different task. Nevertheless, this finding raises the question; does recollection automatically occur during comparative judgments? Theoretically all the information necessary to make decisions is available in an earlier familiarity process, and a truly smart memory system would terminate any further unnecessary search or recollection.

A closer examination of the task sequence might help explain why Rosburg et al. (2011) found significant parietal/old new effects during RH trials and we did not. In order to capture ERPs during the city-size comparison task, it was necessary to show each region within a pair sequentially. In RH trials where the first region is recognized, that region should elicit some positive familiarity signal. It is also likely, because the participants are not yet aware that the second region will be unknown to them, that they would engage in some form of recollection or memory search for cues that might help them in their decision. Once the second region is displayed, and indeed unknown to the participant, there should be no familiarity and no recollection. In RH trials where the first region is unknown, there is again no positive amount of familiarity and no information available to recollect. However, the participant now knows that any type of quick recognition of the second region will likely suffice to make a decision. The second region appears, and is associated with some positive familiarity signal that can be utilized to choose that region as more populous. Further knowledge is not necessary for the decision, and it is possible that our memory system recognizes this fact and terminates recollection before it occurs. The end result, assuming equal amounts of recognized-first and unrecognized-first trials, is approximately 100% of trials with a reliable familiarity difference between the two regions,
and approximately 50% of trials with a reliable recollection difference between the two regions, assuming recollection terminates in the unrecognized-first sequence. If true, this would seem to imply enhanced and conditional control of memory during decision making, such that memory processes would only be engaged insofar as they are valuable to the decision process. The difference between Rosburg et al.’s (2011) significant results and our own non-significant results concerning parietal old/new effects could be attributable to two sources. Firstly, Rosburg et al. paired the regions in their task based on a-priori recognition rates, meaning that in a number of trials, both regions were likely recognized by participants and therefore were not RH trials, but FH trials instead. It is possible that recollection processes were recruited to help make decisions during these trials. Secondly, Rosburg et al. chose an earlier time window (450 - 600 ms) when statistically evaluating their parietal old/new effects compared to our time window (500 - 800 ms). A closer evaluation of Rosburg et al.’s waveforms shows that by 800 ms, the parietal old/new effect difference has extinguished, and the waveforms have actually begun to reverse polarities. It is likely that by incorporating the latter tail of the parietal old/new effects in their analysis, the authors might have found non-significant differences between recognized and unrecognized regions. Both Rosburg et al.’s results and our own show a more robust FN400 familiarity difference for RH trials than parietal old/new effect recollection differences, perhaps reflective of recollection differences occurring in only 50% of the captured trials, when the recognized region was presented first. This is a question that warrants further examination in the future and can be investigated by a re-analysis of the current data. Regardless of the presence of recollection during RH trials, it is clear that information is available in early familiarity processes to make reliable decisions, and that participants likely rely on this familiarity (or recognition) cue. In RH cases, we are in agreement with Hilbig et al.
(2011) that familiarity is driving population decisions. However, our current results lead us to posit a role for recollection, not familiarity, in FH cases. Hilbig et al. claim that the same underlying familiarity processes guiding RH decisions are guiding FH decisions, the only difference being that FH cases hinge on a smaller memory range because it is contingent upon positive amounts of recognition, and thus necessitates at least some intermediate degree of familiarity for both regions within a pair. Although the authors seemingly contend there is still exploitable information in the comparison of these intermediary familiarity signals, our results contradict this. ERPs during FH trials from Experiment 1 showed no distinguishable FN400 familiarity effects between more quickly and less quickly retrieved regions within a pair, seemingly indicating that the two familiarity signals, at least as measured by the FN400, would not provide robust enough information to dissociate the two regions. Furthermore, these trials were associated with large parietal old/new effect differences, indicating greater recollection occurring for more quickly retrieved regions. Not only does a comparison of familiarity signals in these intermediary situations provide less robust information, this information is likely less valid compared to the information provided by recollection processes. Recollection allows an influx of potentially relevant knowledge that could contain valid cues, such as whether or not a region has an airport or a professional sports team. This information is not only more valid than a coarse familiarity signal, but also likely more valid than retrieval fluency, and it would be unwise for a decision maker to ignore this information in situations where accuracy is emphasized.

Hilbig et al. (2011) conclude that once the recognition cue becomes weighted by familiarity, it can incorporate all RH and FH trial types, and there is no need for an additional FH mechanism that emphasizes use of retrieval fluency (a secondary product of familiarity). Although we agree there are problems in the way the RH and FH have been conceptualized, it
still appears that two different cues are being relied upon in different situations. We are not ready to entirely abandon the necessity of a second FH-like mechanism, because, “having two distinct rules improves the overall efficiency of the system because information is processed only as much as is necessary to make a decision” (Schooler & Hertwig, 2005, p. 626). To put it simply, our findings seem to suggest that familiarity can act as an early, relatively effortless screening signal used to dissociate two objects, as long as the difference between the two familiarity signals is large enough to be exploited. This interpretation of the operation of the RH and FH is consistent with early literature on recognition memory put forth by Atkinson and Juola (1973, 1974). This work asserted that when making recognition judgments, participants would either (a) make an initial fast response based on familiarity or (b) if the familiarity is neither high nor low, delay responding until an extended search of memory is carried out. Atkinson and Juola’s work on recognition corroborates our assertion that familiarity is only utilized to make decisions when there is a large disparity in the familiarity of two regions. These cases often correspond to RH cases due to the conditional necessity that one object be unknown. However, it still seems possible that two recognized regions - one with a very faint familiarity signal and one with a strong familiarity signal, could indeed be dissociated based on the familiarity cue. This scenario would violate the current definition of the RH, but is consistent with Hilbig et al.’s (2011) perspective. Once the difference between the familiarity signals of two objects becomes indistinguishable, decision makers can rely on the slightly more effortful recollection process to provide cues about decisions, assuming some type of recollectable knowledge is available for at least one of the objects. This perspective is once again corroborated by Atkinson and Juola’s work on recognition memory, suggesting that participants turn to further memory search and recollection for answers when familiarity does not reliably differentiate two regions. These cases
FAMILIARITY AND RECOLLECTION IN DECISION MAKING

often correspond to FH cases, again due to the conditional nature of the FH, which states that both objects must be associated with some level of recognition. Although pure retrieval fluency still seems to be playing a minor role in these decisions (no more than one-quarter of the times it is applicable), our findings coupled with Hilbig et al.’s provide evidence against a prominent role for use of retrieval fluency in isolation as a decision making cue.

Conclusions

We began this thesis with a strong impression of the RH and FH as discrete decision making mechanisms. A more careful assessment of the underlying memory processes seems to suggest a distinction based less upon the dichotomy and speed of recognition, but rather the strength of memory itself. RH decisions have typically been considered to be made based on mere recognition of one stimulus within a pair, and we have shown that insofar as recognition parallels a positive underlying level of familiarity, this remains true. Because a recognized and unrecognized object will necessarily be associated with disparate levels of familiarity, the RH can capitalize on this information to make a quick, accurate decision with little effort. FH decisions, on the other hand, have typically been presumed to rely on retrieval fluency or speed of recognition to make judgments. Prior research has suggested that an underlying familiarity process also governs this retrieval fluency, upon which decisions are based. However, our research indicates that in situations in which the FH has been presumed to operate, underlying levels of familiarity for two stimuli are often too similar to produce a reliable cue upon which to base decisions. When a reliable familiarity difference between two objects is not accessible, we propose that decision makers turn to a more effortful recollection process to inform their choices on a majority of trials. Our results support the hypothesis that the RH and FH, due to their conditional natures, seem to capture instances where decision makers are indeed relying on
different memory-based cues - familiarity during RH cases and recollection during FH cases. Recognition speed (fluency) during FH cases, however, likely plays an auxiliary role to recollection, and thus causes us to question the practicality of the “fluency heuristic” as it is traditionally defined.
References


### Appendix A

#### Experiment 1: r-s-Model Categories and Observed Choice Frequencies

<table>
<thead>
<tr>
<th>r-s-model category</th>
<th>r-s-model category meaning</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fluency-homogenous FH case, correct judgment</td>
<td>2821</td>
</tr>
<tr>
<td>2</td>
<td>Fluency-homogenous FH case, false judgment</td>
<td>1410</td>
</tr>
<tr>
<td>3</td>
<td>Neither recognized (guess), correct judgment</td>
<td>260</td>
</tr>
<tr>
<td>4</td>
<td>Neither recognized (guess), false judgment</td>
<td>205</td>
</tr>
<tr>
<td>5</td>
<td>One recognized, adherence to RH, correct judgment</td>
<td>3065</td>
</tr>
<tr>
<td>6</td>
<td>One recognized, adherence to RH, false judgment</td>
<td>895</td>
</tr>
<tr>
<td>7</td>
<td>One recognized, non-adherence to RH, false judgment</td>
<td>281</td>
</tr>
<tr>
<td>8</td>
<td>One recognized, non-adherence to RH, correct judgment</td>
<td>157</td>
</tr>
<tr>
<td>9</td>
<td>Fluency-heterogeneous FH case, adherence to FH, correct</td>
<td>4406</td>
</tr>
<tr>
<td></td>
<td>judgment</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Fluency-heterogeneous FH case, adherence to FH, false</td>
<td>1741</td>
</tr>
<tr>
<td></td>
<td>judgment</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Fluency-heterogeneous FH case, non-adherence to FH, false</td>
<td>1586</td>
</tr>
<tr>
<td></td>
<td>judgment</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Fluency-heterogeneous FH case, non-adherence to FH,</td>
<td>2373</td>
</tr>
<tr>
<td></td>
<td>correct judgment</td>
<td></td>
</tr>
</tbody>
</table>

### Appendix B

#### Experiment 2: r-s-Model Categories and Observed Choice Frequencies

<table>
<thead>
<tr>
<th>r-s-model category</th>
<th>r-s-model category meaning</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fluency-homogenous FH case, correct judgment</td>
<td>756</td>
</tr>
<tr>
<td>2</td>
<td>Fluency-homogenous FH case, false judgment</td>
<td>448</td>
</tr>
<tr>
<td>3</td>
<td>Neither recognized (guess), correct judgment</td>
<td>293</td>
</tr>
<tr>
<td>4</td>
<td>Neither recognized (guess), false judgment</td>
<td>284</td>
</tr>
<tr>
<td>5</td>
<td>One recognized, adherence to RH, correct judgment</td>
<td>2337</td>
</tr>
<tr>
<td>6</td>
<td>One recognized, adherence to RH, false judgment</td>
<td>658</td>
</tr>
<tr>
<td>7</td>
<td>One recognized, non-adherence to RH, false judgment</td>
<td>354</td>
</tr>
<tr>
<td>8</td>
<td>One recognized, non-adherence to RH, correct judgment</td>
<td>194</td>
</tr>
<tr>
<td>9</td>
<td>Fluency-heterogeneous FH case, adherence to FH, correct</td>
<td>3620</td>
</tr>
<tr>
<td></td>
<td>judgment</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Fluency-heterogeneous FH case, adherence to FH, false</td>
<td>1490</td>
</tr>
<tr>
<td></td>
<td>judgment</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Fluency-heterogeneous FH case, non-adherence to FH, false</td>
<td>1257</td>
</tr>
<tr>
<td></td>
<td>judgment</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Fluency-heterogeneous FH case, non-adherence to FH,</td>
<td>1692</td>
</tr>
<tr>
<td></td>
<td>correct judgment</td>
<td></td>
</tr>
</tbody>
</table>