

# The Simple Life: New experimental tests of the recognition heuristic

Zachariah Basehore\*

Richard B. Anderson†

## Abstract

The recognition heuristic (RH) is a hypothesized decision strategy that is assumed to enable individuals to make decisions quickly and with minimal effort. To further test this hypothesized strategy, an experiment assessed the proportion of RH-consistent selections when recognition was unconfounded with any other cues (at the group level). This was accomplished by showing participants a fictitious city in the beginning of the experimental procedure, before asking them to decide whether the previously presented city or a novel fictitious city has the larger population. As hypothesized, people made significantly more RH-consistent selections than chance. Thus, Experiment 1 demonstrated that RH can explain a considerable proportion of participant decisions in a procedure that experimentally excluded alternative interpretations of that behavior. In a second experiment, each participant was given a training session with accuracy feedback. In one group, well-known cities were larger on 80% of trials. In another group, well-known cities were larger on 50% of trials. In a third group, well-known cities were larger on only 20% of trials. On a judgment task later in the procedure, on which there was no feedback, participants from the third group made significantly fewer RH-consistent selections than those in the first two groups. Overall, the present results experimentally remove potential confounds and ambiguities that were present in many prior studies. Specifically, Experiment 1 establishes that people's choice of recognized over unrecognized objects truly does reflect the use of recognition, rather than other cues; Experiment 2 experimentally demonstrates that learned recognition validity affects the use of recognition, even with a small training sample.

Keywords: The Simple Life, recognition heuristic, decision making, experimental tests.

## 1 Introduction

Gerd Gigerenzer and colleagues advocate a "fast-and-frugal heuristics" approach to decision-making, which is intended to shed light on how humans make adaptive decisions in the real world of scarce information and substantial time pressure (Gigerenzer & Todd, 1999; Hertwig & Todd, 2003). To this end, researchers in the fast-and-frugal program have sought to delineate heuristics that provide specific, easily-modeled explanations for human decisions.

One of the most fundamental heuristics in this program is the recognition heuristic (RH), proposed by Goldstein and Gigerenzer (1999; 2002). RH is defined: "If one of two objects is recognized and the other is not, then infer that the recognized object has the higher value with respect to the criterion" (Goldstein & Gigerenzer, 2002, p. 76). They also wrote that "If one object is recognized and the other is not, then the inference is determined; no other information about the recognized object is searched for and, therefore,

no other information can reverse the choice determined by recognition" (Goldstein & Gigerenzer, 2002, p. 82). They further noted that this strategy is useful only when recognition is strongly correlated with the criterion (p. 87). Later, research by Pachur and Hertwig (2006) corroborated this notion, showing that the proportion of decisions consistent with RH was notably lower when recognition validity was low (though participants in this study still made more RH-consistent selections than chance).

To illustrate this effect: if a person (who we'll call "Bob") has heard of Seoul, South Korea but not Daegu, South Korea, then Bob will select Seoul as being more populous.

Indeed, the initial experiments conducted by Goldstein and Gigerenzer (1999; 2002) showed that, when participants were asked to select which of a pair of cities is more populous, about 90% of participants' choices were consistent with RH.

Beginning with Oppenheimer (2003), a series of studies challenged RH. Oppenheimer showed that when people were asked whether a small, local city or an unknown (fictitious) city was more populous, the majority of people selected the unknown city. A second experiment replicated this finding with cities that were well-known for reasons other than size (e.g. Chernobyl vs. a fictitious city). Newell and Shanks (2004) demonstrated that, in a stock market simulation, people consider both recognition (which was artificially induced by the experimental procedure) and expert advice. Bröder and Eichler (2006) found that participants

---

The authors thank Carolyn J. Tompsett and Patrick J. Nebl for their assistance in interpreting statistical modeling procedures, as well as the editor and two anonymous reviewers for their helpful comments and recommendations on an earlier draft.

Copyright: © 2016. The authors license this article under the terms of the Creative Commons Attribution 3.0 License.

\*Department of Psychology, Bowling Green State University, Bowling Green, Ohio, 43403, USA. Email: zdbaseh@bgsu.edu.

†Department of Psychology, Bowling Green State University.

considered recognition alongside other cues, such as the presence of a soccer team or an intercity train line, when deciding which of two cities is more populous. Richter and Späth (2006) and Pohl (2006) also found that recognition is considered as just one cue among many. Hilbig (2010) reviewed this and other evidence in a wide-ranging paper that discussed limitations in the evidence for the fast-and-frugal program.

Pachur, Bröder and Marewski (2008) found evidence that participants often do *not* override recognition when making judgments — even when the participants themselves judged the presence or absence of an international airport to be a more valid cue than recognition. These authors suggested that participants behave differently when they rely on knowledge from outside the laboratory setting, rather than cues presented within the experimental procedure. Pachur et al. (2011) further explained that RH is applicable particularly under conditions of uncertainty, rather than when more conclusive knowledge is available (p. 2).

In a consumer decision-making task, Oeusoonthornwattana and Shanks (2010) showed limited support for RH. Participants chose a product by the recognized brand on more than 50% of trials, though further information about the company's behavior also influenced participants' selections. Pachur and Hertwig (2006) suggested that people “evaluate” or “filter” RH before applying it (p. 999), as did Newell and Shanks (2004) and Newell (2011).

Most of these existing studies used the proportion of RH-consistent selections (sometimes called “RH-adherence” — a term that presupposes that people use RH) as a measure of whether RH was used on a given trial. However, as Hilbig (2010) notes, it is problematic to use the *outcome* to measure the use of a psychological *process*, as multiple different processes could lead to the same outcome. Hilbig, Erdfelder and Pohl (2010) used a statistical modeling procedure to separate the effect of recognition from that of other information. They found that participants appeared to use recognition to a substantial extent, though the model indicated that participants also considered further valid information. Furthermore, they classified 55% of participants (in their last data set) as non-users of RH, despite an RH adherence rate in this group that was nearly the same as the adherence rate among those who were classified as RH users. This study shows the value of statistical modeling to isolate recognition from the effects of other information. It did not address whether such isolation may be accomplished by means of experimental control.

The apparent suspension of RH is another area where the findings are not in agreement. Several studies (e.g. Newell & Shanks, 2004; Pachur & Hertwig, 2006; Bröder & Eichler, 2006; Hilbig, Erdfelder & Pohl, 2010; Horn, Pachur & Mata, 2015) found that further knowledge can adjust, or even completely override, the effect of differential recognition on participants' decisions. But Pachur, Bröder &

Marewski (2008) found the opposite: Though participants learned further cues that contradicted recognition, and explicitly rated some of the contradictory cues as being *more* valid than recognition, participants made RH-consistent selections on over 80% of applicable trials (on average) — even when three further cues contradicted recognition.

As Goldstein and Gigerenzer (2002, p. 87) noted, RH is helpful only when recognition validity is higher than chance. Hogarth and Karelaia (2005), and Davis-Stober, Dana and Budescu (2010), among others, used mathematical modeling and simulations to show that when a single cue is highly correlated with the criterion, a single-cue model can outperform information-integration strategies at generalizing to future data that has a high degree of uncertainty. Czerlinski, Gigerenzer & Goldstein (1999) showed that this effect holds with a variety of real-world data as well. So, it follows that participants should decrease their reliance on differential recognition when recognition validity is low, as in Pachur and Hertwig (2006). However, people could use low recognition validity itself as a cue, as in Experiment 1 of Richter and Späth (2006) — the knowledge that the Siberian tiger is an endangered species automatically indicates that few Siberian tigers are currently alive.

But how do people know when recognition validity is high enough to warrant a strategy like RH? And how much training is required before people learn whether or not a certain strategy is appropriate? An obvious possibility is that each individual evaluates a simple strategy like RH based on feedback gained from life experience. If the strategy is often incorrect, then it is likely to be adjusted, or discarded completely. But if it is often effective, a simple strategy will continue to be implemented in those domains in which it has proven useful.

Bröder and Eichler (2006) found that participants' use of recognition varied with its validity, when they provided participants with explicitly stated validities of four different cues to city population, including recognition. The results also showed that participants' selections incorporated all available information, including (but not limited to) recognition. In a study using fictitious company names, Newell and Shanks (2004) used a similar approach, and found similar results.

Pachur and Hertwig (2006) found that when people judged the incidence rate of infectious diseases, the proportion of RH-consistent selections was 62%; much lower than the 90% in Goldstein and Gigerenzer's (2002) original city task. This study indicates that people's use of recognition varies with recognition validity, and that people attend to a cue's validity in everyday life, not just when cue validities are explicitly provided in a laboratory setting.

Pohl (2006; Experiment 1) compared judgments of a Swiss city's population to judgments of a Swiss city's distance from the geographical center of the country. In this study, recognition was uncorrelated with the distance of a

city from the geographical center of Switzerland, but was highly correlated with the city's population. Participants often made decisions consistent with RH on the population judgment task (89% of trials), but picked the recognized city on only 54% of trials in the distance judgment task. In Experiment 2, Pohl (2006) tested Swiss cities (some of which were well-known and large) against Swiss towns with ski resorts (which are often small, even when they *are* well-known). Participants made selections consistent with RH on 86% of trials in which it led to the correct inference, but on 46% of trials in which recognition led to the incorrect inference. Both of those experiments also show that RH is sensitive to recognition validity.

Existing studies have not manipulated recognition validity above and below chance level within a single experiment in which the judgment was held constant (e.g., which of two cities had the larger population), and thus have not systematically examined the effect of recognition being positively correlated vs. uncorrelated vs. negatively correlated with the criterion. A controlled study examining this effect can help to more precisely determine the effect of recognition validity on RH use.

## 2 Experiment 1

To accomplish the goal of separating recognition from further knowledge, participants in Experiment 1 were given prior exposure to one set of fictitious city names, but no pre-exposure to another set of fictitious city names. Participants were later asked to choose which of two cities was more populous; one of the pair had been pre-exposed earlier in the experiment (the identity of the pre-exposed city was counterbalanced across participants); the other city name was novel. Therefore, this procedure employs experimental control to divorce recognition from other salient information on which participants could base a decision. Because the assignment of city names to conditions is counterbalanced across participants, this eliminates any potential confound at the group level between the city names' particular linguistic (or other) characteristics, and whether or not they were pre-exposed.

Consistent with most prior research, participants were expected to be more likely to select the recognized city over an unrecognized one. But what if people use recognition only when further knowledge is *also* available? It is not evident that prior studies have ruled out this possibility, though Goldstein and Gigerenzer (2002; Study 3) and Hilbig, Erdfelder and Pohl (2010) have provided the best attempts thus far. In light of existing research, it is possible that people do not consider recognition at all; rather, they may rely upon the confluence of recognition with other cues. If either possibility is the case, then participants in the present experiment would guess randomly, picking the pre-exposed city

no more often than chance. The present experiment therefore provides a clear, falsifiable test of the assumption that mere recognition is used as a decision cue.

Another advantage of the present procedure over prior studies is that it does not rely on participants' self-reported recognition (though such self-reports will be measured and incorporated into some of the analyses). Many prior studies have required participants to undergo hundreds of trials, viewing dozens of different cities. It is very possible that cognitive fatigue sets in by the end of the experiment — which is usually when participants are asked to identify which cities they recognized. If participants made a mistake in their reporting, the analysis of RH-consistent selections may not be entirely accurate. Since the present procedure uses tightly controlled stimuli and only ten trials per participant, limitations such as cognitive fatigue or reporting errors can be circumvented.

### 2.1 Method

**Participants.** Thirty students at Bowling Green State University agreed to participate in Experiment 1, in partial fulfillment of course requirements (16 women and 14 men, ages 18–23, mean age 19.21).

**Procedure.** Using Qualtrics web-based survey software, all participants completed this experiment on a computer in a laboratory at Bowling Green State University. Participants were run one at a time.

Twenty fictitious cities were used for this experiment (see Table 1). Some were originally used by Oppenheimer (2003); others were invented by the primary author. In the first stage (the pre-exposure phase), participants were asked whether or not they recognized a fictitious city. To further encourage encoding of the stimulus, they were also asked to indicate their degree of confidence in their judgment. This procedure was repeated ten times, with ten different fictitious cities, for the purpose of inducing recognition for those cities.

Next, in the judgment phase, each city from the pre-exposure phase was paired with a novel fictitious city that was purportedly from the same country as the pre-exposed city (the order of response options in each pair was counterbalanced across participants). Participants were asked to select the city they thought was more populous. On the same web page, a manipulation check required participants to indicate whether they had seen each of the cities from outside the experiment, from earlier in the experiment, or never before. This allowed for an evaluation of RH-consistent selections on an individual-by-individual basis.

For instance, participants were asked if they recognized Weingshe, China (among nine other fictitious cities, in random sequence) in the pre-exposure phase. In the judgment phase, participants were asked whether they thought Weing-

Table 1: Fictitious city names in Experiment 1.

Pre-exposure Phase and Judgment Phase*	Judgment Phase Only*
<i>Heingjing</i> , China	Huanlizhen, China
Nehaiva, Israel	<i>Gohaiza</i> , Israel
<i>al-Ahbahib</i> , United Arab Emirates	al-Fashik, United Arab Emirates
<i>Papayito</i> , Mexico	<i>Las Besas</i> , Mexico
<i>Weingshe</i> , China	Meingzhao; China
<i>Rhavadrán</i> , India	Vedharal, India
<i>Schretzburg</i> , Austria	Nyksbörg, Austria
Ohigmoso, Slovakia	Ramslinn, Slovakia
Bórszki, Poland	Waszlów, Poland
Áventyrnisse, Iceland	Thaskörvik, Iceland

Cities on the same row were presented as a pair in the judgment phase. Italics denote names originally used by Oppenheimer (2003).

\*The assignment of cities to phases (the left column vs. the right column in the table) was counterbalanced across participants.

she, China or Meingzhao, China was more populous (among nine other city pairs, again presented in random sequence), also indicating whether they had seen each city outside the experiment, earlier in the experiment, or never before.

The assignment of cities to the pre-exposure phase was counterbalanced across participants. Thus, half of the participants (15 of 30) saw the stimulus set as shown in Table 1. For the other half, the columns were swapped. For example, some participants saw Weingshe, China in the pre-exposure phase before being asked in the judgment phase whether Weingshe or Meingzhao is more populous. Other participants saw Meingzhao in the pre-exposure phase, before being asked to choose between Weingshe and Meingzhao.

## 2.2 Results

**RH-consistency.** In one set of analyses — consistent with the procedures of previous studies such as Goldstein and Gigerenzer (2002), Oppenheimer (2003), Newell and Shanks (2004), Pachur and Hertwig (2006), and Bröder and Eichler (2006) — we excluded judgment trials on which the participant reported recognizing neither city or both cities in a pair, since RH would not be applicable on such trials. This resulted in the exclusion of 28% of the trials.

A one-sample  $t$  test showed that participants' mean proportion of RH-consistent selections was significantly greater than chance ( $M = 0.743$ ,  $SD = 0.212$ ,  $t(29) = 6.27$ ,  $p < .001$ , 95% CI [0.664, 0.823]). Cohen's  $d$  indicates that this is a large effect size ( $d = 1.15$ ). Additionally, to assess whether the findings were influenced by the assignment of stimuli to particular roles within the procedure, we conducted a one-factor, two-condition ANOVA. Stimulus assignment (A or

B; Condition A indicates the ordering shown in Table 1; Condition B swapped the contents of the columns in Table 1) was the independent variable, and RH-consistency was the dependent variable. No significant effect was found, ( $F(1,28) = 0.440$ ,  $p = .512$ ). Additionally, one-sample  $t$  tests for each of the two stimulus assignments showed that the mean proportion of RH-consistent selections was significantly greater than chance for Condition A ( $M = 0.769$ ,  $SD = 0.214$ ,  $t(14) = 4.863$ ,  $p < .001$ , Cohen's  $d = 1.26$ ) as well as for Condition B ( $M = 0.717$ ,  $SD = 0.215$ ,  $t(14) = 3.922$ ,  $p = .002$ , Cohen's  $d = 1.01$ ).

In order to rule out the possibility that participants were systematically using a cue other than recognition, we also conducted tests in which the unit of analysis was city pair, rather than participant. The results were consistent with the prior analyses. RH-consistency was significantly greater than chance ( $M = 0.741$ ,  $SD = 0.134$ ,  $t(19) = 8.015$ ,  $p < .001$ , Cohen's  $d = 1.80$ ). This was a group-level analysis only, as there was no way to determine whether any idiosyncratic individual variations in cue usage existed.

**Pre-exposure consistency.** Since recognition was experimentally controlled, it is not necessary to rely on participants' self-reported recognition. An analysis of the proportion of times the pre-exposed city was chosen (regardless of the applicability of RH based on self-reported recognition) showed that the pre-exposed city was selected significantly more often than chance ( $M = 0.68$ ,  $SD = 0.207$ ,  $t(29) = 4.75$ ,  $p < .001$ , 95% CI [0.603, 0.758]). Cohen's  $d$  shows a large effect size ( $d = 0.868$ ).

As with RH-consistency, we also conducted a test in which the unit of analysis was city pair, rather than partici-

pant. The results indicated that the mean proportion of pre-exposure-consistent choices was significantly greater than chance ( $M = 0.65$ ,  $SD = 0.151$ ,  $t(19) = 4.436$ ,  $p < .001$ , Cohen's  $d = 0.99$ ).

**RH-consistency vs. Pre-exposure consistency.** Why the 6 percentage point difference between RH-consistency and pre-exposure consistency? A higher proportion of decisions were consistent with RH (74.3%) than were consistent with pre-exposure (68.0%). To evaluate whether this difference is significantly different from zero, we subtracted (for each participant) the proportion of decisions that were consistent with pre-exposure from the proportion of decisions that were consistent with recognition. A one-sample  $t$  test showed that this difference is significantly different from zero ( $M = 0.063$ ,  $SD = 0.097$ ,  $t(29) = 3.589$ ,  $p = .001$ , 95% CI [0.027, 0.099], Cohen's  $d = 0.65$ ).

This pattern is due to the inclusion of the RH-inapplicable trials in the analysis of pre-exposure consistency. Since 28% of trials were excluded from the analysis of RH consistency because RH did not apply (participants reported recognizing neither city, or both cities), and due to the fact that these were fictitious cities with no salient cues other than recognition, participants would have been forced to guess on those trials. Indeed, among trials on which RH did not apply, participants selected the pre-exposed city 49.6% of the time, indicating that they truly were guessing when RH could not be used.

**Individual data.** Gigerenzer and Goldstein (2011) called for researchers to report individual data, not just aggregate means. Consistent with that call, an individual analysis showed that 8 out of 30 participants made RH-consistent decisions on every trial; 2 more made RH-consistent decisions on 90% of trials. Approximately one-third of participants might therefore be classified as RH users.

Again, the data shows a lower proportion of judgments consistent with pre-exposure than with RH: 5 out of 30 participants made decisions consistent with pre-exposure on every trial; 3 more made such judgments on 90% of trials.

## 2.3 Discussion

A single exposure to a stimulus was enough to bias participants to choose the pre-exposed option — even though recognition was induced artificially by the experimental procedure itself. That the proportion of RH-consistent selections was significantly higher than chance indicates that recognition is indeed a significant factor; the large effect size speaks to the reliability of this effect.

Numerous studies have already demonstrated that people make significantly more RH-consistent selections than chance, notably Goldstein and Gigerenzer (2002), Newell and Shanks (2004), Pachur and Hertwig (2006), Bröder and Eichler (2006), Pachur, Bröder and Marewski (2008), Marewski et al. (2010), Oeusoonthornwattana and Shanks

(2010), Hilbig, Erdfelder and Pohl (2010), and Horn, Pachur and Mata (2015).

However, in these studies, at least one cue other than recognition was present among the stimuli, such as the presence of an international airport (Pachur, Bröder & Marewski, 2008), number of mentions in popular media as a cue to the incidence rate of a disease (Pachur & Hertwig, 2006), any further knowledge about real cities that participants may have brought from their life experiences (Goldstein & Gigerenzer, 2002; Bröder & Eichler, 2006;<sup>1</sup> Marewski et al., 2010; Hilbig, Erdfelder & Pohl, 2010; Horn, Pachur & Mata, 2015), recommendations from fictitious “experts” (Newell & Shanks, 2004), or positively/negatively valenced information about a company (Oeusoonthornwattana & Shanks, 2010). The apparent use of recognition in those studies may have therefore been a result of the integration of several cues, rather than reflecting the use of recognition alone, as Hilbig (2010) argues. The results presented here confirm that participants do use recognition as a basis for decisions, to a significant extent.

Goldstein and Gigerenzer (2002, p. 85) found evidence for a type of pre-exposure effect in which the repetition of real, previously unrecognized cities increased the likelihood of those cities being judged larger than other real cities that had not been repeated. That study, like the present one, did not provide participants with additional, non-recognition information. However, the present findings make a number of additional contributions. First, they provide a replication of Goldstein and Gigerenzer's study. Second, the design of the present study, which counterbalanced the assignment of fictitious city names to roles within the experiment, ensured that nothing about the city names could have influenced participants' choices. Third, the present procedure and analysis method permitted the assessment of both pre-exposure *and* RH-consistency (computed only for those trials to which RH could apply) for the same data set.

As mentioned earlier, Hilbig, Erdfelder and Pohl's (2010) multinomial processing tree model determined the probability that participants were relying on recognition. Whereas Hilbig et al. (2010) used a statistical modeling approach, the present study employed the power of experimental control to provide converging evidence for the importance of recognition in the decision-making process.

The findings from the present methodology replicate and extend a variety of prior research by vividly demonstrating that recognition does form a robust basis for human decision-making, even when the experimental stimuli are very tightly controlled. Along these lines, Pachur and Hertwig (2006) have already made a persuasive case that recognition is “first on the mental stage” and therefore provides a significant initial bias for people's decisions (p. 986). The present experiments confirm the importance of recog-

<sup>1</sup>Arguably, the cities used by Bröder & Eichler were totally unfamiliar, but they were real and some may have been recognized.

nition as a decision cue. However, the present Experiment 1 yielded RH-consistent selections on fewer than three-quarters of trials, a finding that is consistent with several studies that have shown much lower proportions of RH-consistent selections than the original 90% found by Goldstein and Gigerenzer (2002; Study 1).

Several previous studies (e.g., Oppenheimer, 2003; Newell & Shanks, 2004; Bröder & Eichler, 2006; Hilbig, Erdfelder & Pohl, 2010; Oeusoonthornwattana & Shanks, 2010) have demonstrated that people do consider cues other than recognition, contrary to Goldstein and Gigerenzer's (2002) claim that "...no other information can reverse the choice determined by recognition" (p. 82). In that context, these results show the limitations of using mere recognition to explain choices.

Experiment 1 does not, however, provide evidence as to the circumstances under which people might modify their use of recognition. A second experiment addresses this question.

### 3 Experiment 2

In Experiment 2, a training phase with immediate and salient feedback was followed by a judgment task (with no feedback). The training phase was intended to provide information about recognition validity, which participants could use to inform their strategy in the subsequent judgment task. Depending on the training condition to which a participant was assigned, recognition validity was positively correlated, uncorrelated, or negatively correlated with the correct choice. Three different training conditions were used: positive recognition validity, zero recognition validity, and negative recognition validity.

Some previous studies have attempted to test the proportion of choices consistent with RH when recognition validity is negative (e.g., Richter & Späth, 2006, Experiment 1). Richter and Späth found that both recognition and further knowledge about whether a species is endangered contributed to the proportion of correct responses. In other studies, such as McCloy, Beaman, Frosch and Goddard (2010) and Hilbig, Scholl and Pohl (2010), when participants were asked to choose the *smaller* of two cities, the proportion of RH-consistent decisions decreased compared to when participants were asked to choose the *larger* of two cities. However, no prior studies have systematically altered participants' perceptions of their own recognition validity across the positive, zero, and negative validities for stimuli from the same reference class. By doing this, the present procedure allows for a direct comparison of the effect of different levels of recognition validity on participants' selections.

Participants in the negative recognition validity training condition were expected to make fewer RH-consistent selections on the subsequent judgment task than participants

in the other two groups; participants in the zero recognition validity training were expected to make slightly fewer RH-consistent selections than participants in the positive recognition validity condition.

#### 3.1 Method

**Participants.** Seventy-seven Bowling Green State University students participated in Experiment 2, in partial fulfillment of course requirements.

The data for six of these participants were excluded because they reported that they never, or only once, recognized one of the two alternatives. In the former case, the participant's proportion of RH-consistent responses was undefined, because the proportion is computed only for those trials on which there is differential recognition. In the latter case, the participant's data were binary: if differential recognition was applicable on only a single trial, that participant made either one RH-consistent response (yielding 100% RH-consistency) or zero RH-consistent responses (yielding 0% RH-consistency). Such a binary outcome would be unsuitable for the analysis employed here.

This left 71 participants in the final analysis; 50 women, 20 men, and 1 participant who identified as "other" ages 18–42 ( $M = 19.49$  years,  $SD = 3.08$ ).

**Procedure.** Unlike the first experiment, Experiment 2 used real cities to test the effect of learning with feedback on the proportion of RH-consistent selections participants made on a subsequent judgment task. The computer software and laboratory setting were the same as in Experiment 1.

Each participant was first given a training set. In the training set, the participant was asked to judge which of two cities (matched by country) was more populous (Table 2). After each trial, participants received feedback on whether the response was correct or incorrect, as well as the population of both cities in that trial. At the end of the training set, participants were also shown their overall accuracy.

Participants were randomly assigned to one of 3 training sets, with each set having a different level of cue validity. There was the .8 cue-validity condition (in which the correlation between the cue and the criterion was 0.6), the .5 condition (in which the cue and criterion were uncorrelated), and the .2 condition (a  $-.6$  correlation between the cue and the criterion).

Because the stimuli were real city names, cities were selected for each training set based on the results of a pilot study. For the pilot study, fifty people reported which cities they recognized, from a list of 100 cities drawn from a demographic database (Demographia, 2014). After tallying the number of people who recognized each city, the 31 pairs of cities (matched by country) that yielded the greatest differential rates of recognition were selected for the experi-

Table 2: City names used in Experiment 2.

Training			Judgment
Low Validity (0.2)	Medium Validity (0.5)	High Validity (0.8)	
<i>Paris, France*</i>	<i>Paris, France*</i>	<i>Paris, France*</i>	<i>Vancouver, Canada*</i>
Le Havre, France	Le Havre, France	Le Havre, France	Kelowna, Canada
<i>Istanbul, Turkey*</i>	<i>Istanbul, Turkey*</i>	<i>St. Petersburg, Russia*</i>	<i>Buenos Aires, Argentina*</i>
Izmir, Turkey	Izmir, Turkey	Ekaterinburg, Russia	Salta, Argentina
Hong Kong, China*	<i>Dubai, United Arab Emirates*</i>	<i>Rome, Italy*</i>	<i>Athens, Greece*</i>
<i>Shenzhen, China</i>	al-Ain, United Arab Emirates	Catania, Italy	Thessaloniki, Greece
Cancun, Mexico*	<i>Rio de Janeiro, Brazil*</i>	<i>Dubai, United Arab Emirates*</i>	<i>Munich, Germany*</i>
<i>Queretero, Mexico</i>	Maraba, Brazil	al-Ain, United Arab Emirates	Bremen, Germany
Jerusalem, Israel*	<i>Sacramento, California, USA*</i>	<i>Rio de Janeiro, Brazil*</i>	<i>Tapei, Taiwan*</i>
<i>Haifa, Israel</i>	Fairfield, California, USA	Maraba, Brazil	Tainan, Taiwan
Bristol, United Kingdom*	Hong Kong, China*	<i>Sacramento, California, USA*</i>	<i>Warsaw, Poland*</i>
<i>Tyneside, United Kingdom</i>	<i>Shenzhen, China</i>	Fairfield, California, USA	Lodz, Poland
Fuji, Japan*	Cancun, Mexico*	<i>Mumbai, India*</i>	<i>Brisbane, Australia*</i>
<i>Nagoya, Japan</i>	<i>Queretero, Mexico</i>	Allahabad, India	Hobart, Australia
Acapulco, Mexico*	Jerusalem, Israel*	<i>Melbourne, Australia*</i>	<i>Santiago, Chile*</i>
<i>Torreón, Mexico</i>	<i>Haifa, Israel</i>	Sunshine Coast, Australia	Valparaíso, Chile
Damascus, Syria*	Bristol, United Kingdom*	Hong Kong, China*	<i>Milan, Italy*</i>
<i>Aleppo, Syria</i>	<i>Tyneside, United Kingdom</i>	<i>Shenzhen, China</i>	Palermo, Italy
Quebec City, Canada*	Fuji, Japan*	Jerusalem, Israel*	<i>Tehran, Iran*</i>
<i>Edmonton, Canada</i>	<i>Nagoya, Japan</i>	<i>Haifa, Israel</i>	Rasht, Iran

Cities marked with an asterisk (\*) were more frequently recognized in a pilot study. Those in italics were more populous, according to Demographia (2014).

ment. The city pairs were then selected for each of the training sets, so as to produce the specified cue validities.

In the .8 validity condition, the most-recognized cities (as determined by the pilot study) were more populous on 8 of the 10 trials. In the .5 validity condition, the most-recognized cities were more populous on half of the trials, so recognition was uncorrelated with the criterion. And in the .2 validity condition, the most-recognized cities were more populous on only 2 of the 10 trials. Therefore, this procedure systematically altered participants' perceptions of how useful their own sense of recognition would be as a cue to a city's population.

After completing the training set, all participants performed an experimental judgment task, which consisted of the same stimuli for all participants (regardless of training condition). Participants received no feedback on their accuracy in the judgment task. After this task, participants completed a recognition survey, to facilitate individual analysis of the proportion of RH-consistent selections.

The 0.2 training condition was critical, as it was meant

to train people that recognition would typically lead to the wrong answer (given that the correlation between recognition and the criterion was negative). If participants in the 0.2 condition made fewer RH-consistent choices in the subsequent judgment task than participants in the 0.5 and 0.8 conditions, then this would indicate that negative cue-validities are especially potent in leading people to apparently suspend RH when it is a poor predictor of the criterion.

### 3.2 Results

Again, the analysis was restricted to only the trials on which RH could apply, based on participants' self-reported recognition. RH was not applicable on 41.3% of trials.

A one-way ANOVA showed a significant main effect of treatment condition on the mean proportion of RH-consistent selections ( $F(2, 68) = 5.83, p = .005$ ). Planned-comparison tests of between-group differences revealed that the mean proportion of RH-consistent selections was lower in the 0.2 condition ( $M = 0.829, SD = 0.172, CI[0.755,$

0.904]) compared to the 0.5 ( $M = 0.922$ ,  $SD = 0.105$ ,  $CI[0.878, 0.966]$ ) and 0.8 conditions ( $M = 0.954$ ,  $SD = 0.104$ ,  $CI[0.910, 0.998]$ ). The difference was statistically significant for the 0.2 to 0.5 comparison ( $p = .017$ ) and the 0.2 to 0.8 comparison ( $p = .002$ ), but not for the 0.5 to 0.8 comparison ( $p = .395$ ). A hierarchical regression analysis indicated a positive linear trend (Pearson's  $r(69) = .368$ ,  $p = .002$ ,  $r^2 = .135$ ), but adding a quadratic term did not produce a significant increase in the percentage of variance accounted for ( $r^2$  change = .011,  $F$  change = .867,  $p = .355$ ). Thus, the regression analysis did not support a conclusion that the difference between the judgments of participants in the 0.2 and 0.5 training conditions was significantly different from the 0.5 to 0.8 comparison.

Cohen's  $d$  was again used to calculate effect size. The comparison showed a medium effect size for the 0.2 to 0.5 comparison ( $d = .65$ ) and a large effect for the 0.2 to 0.8 comparison ( $d = .88$ ), but only a small effect for the 0.5 to 0.8 conditions ( $d = .31$ ).

### 3.3 Discussion

Experiment 2 used an experimental methodology, rather than a statistical modeling approach (e.g., Davis-Stober, Dana & Budescu, 2010) to replicate the finding that recognition validity matters, even when training is not extensive and the training set size is small. Feedback from prior experience within the procedure influenced participants' decision strategy. Specifically, the mean proportion of RH-consistent selections was significantly lower for participants who had experienced the .2 training validity condition (in which recognition was negatively correlated with the size of the city) than for participants who had experienced the .5 and the .8 conditions. Thus, the data provide further evidence that cue validities — even those learned over a short period of time — affect people's use of recognition as a decision cue. This finding is important because, as some researchers have argued, training sets are often small in real-world situations (Hertwig & Todd, 2003; Davis-Stober, Dana & Budescu, 2010).

These results extend previous findings regarding recognition validity. Newell and Shanks (2004), Bröder and Eichler (2006), Pachur and Hertwig (2006), and Pohl (2006) all found evidence that the use of recognition is sensitive to recognition validity in a given environment.

A critical point here is that the present Experiment 2 used real cities, about which participants surely had experience from outside the laboratory. It is possible that participants discounted information provided in the experiment itself, compared to what they knew from outside the laboratory. This may explain why the mean proportion of RH-consistent selections was quite high — above .8 — despite those participants having been trained under conditions in which unrecognized cities tended to be larger than recognized ones. As

Pachur and Hertwig (2006) argued, people seem to behave differently with regard to stimuli introduced in the experiment, versus stimuli with which they have prior experience. A reasonable assumption is that this occurs because controlled, artificial stimuli strip out the covariation found in nature, as Brunswik (1955) observed. Future research may be helpful to gain more insight into this phenomenon.

## 4 General discussion

Taken together, these results eliminate any doubt that may remain regarding the powerful role of recognition in the decision-making process, and they also provide evidence that use of a recognition-based strategy is modified based on salient feedback, even from a small training set. Both results also replicate existing findings, which increases our confidence in prior results and reduces the lack of replication in psychology research (Open Science Collaboration, 2015).

Experiment 1 provides evidence that people are more likely to select a recognized option over an unrecognized one. Thus, it seems that RH-consistent selections reflect at least some use of recognition (as opposed to another cue that might be confounded with recognition).

The high rates of RH-consistent selections in Experiment 2, compared to those of Experiment 1, suggest that RH-adherence rates — which are often used to evaluate whether or not a participant is using RH — tend to overestimate the actual use of recognition for natural (as opposed to fictitious) stimuli (see Hilbig, Erdfelder & Pohl, 2010). The difference between RH-adherence rates and pre-exposure adherence rates in Experiment 1 provide further experimental support for this notion. This difference likely occurs because of the covariation between cues that occur in nature, as reported by Brunswik (1955). This adds to previous findings showing that people do incorporate further information, if available, before arriving at a final decision, as a number of previous studies have also found (e.g. Newell & Shanks, 2004; Pachur & Hertwig, 2006; Bröder & Eichler, 2006; Richter & Späth, 2006; Hilbig, Erdfelder & Pohl, 2010; Oeusoonthornwattana & Shanks, 2010).

Although the stimuli (and, to a lesser extent, the methodology) were different between Experiment 1 and Experiment 2, the contrast showcases the wide variety of results obtained by different researchers studying RH. These results reinforce the importance of properly isolating a particular cue, in order to study its impact on decision-making.

## References

- Bröder, A., & Eichler, A. (2006). The use of recognition information and additional cues in inferences from mem-



- ory. *Acta Psychologica*, 121, 275–284. <http://dx.doi.org/10.1016/j.actpsy.2005.07.001>
- Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, 62, 193–217.
- Czerlinski, J., Gigerenzer, G., & Goldstein, D. G. (1999). How good are simple heuristics? In Gigerenzer, Todd, and the ABC Research Group, *Simple Heuristics That Make Us Smart* (p. 97–118). New York, NY: Oxford University Press.
- Davis-Stober, C. P., Dana, J., & Budescu, D. V. (2010). Why recognition is rational: Optimality results on single-variable decision rules. *Judgment and Decision Making*, 5, 216–229.
- Demographia (2014). Demographia world urban areas, 10<sup>th</sup> annual edition. <http://demographia.com/db-worldua.pdf>
- Gigerenzer, G., & Goldstein, D. G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making*, 6, 100–121.
- Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In Gigerenzer, Todd, and the ABC Research Group, *Simple Heuristics That Make Us Smart* (pp. 3–34). New York, NY: Oxford University Press.
- Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In Gigerenzer, Todd, and the ABC Research Group, *Simple Heuristics That Make Us Smart* (p. 37–58). New York, NY: Oxford University Press.
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109, 75–90. <http://dx.doi.org/10.1037//0033-295X.109.1.75>
- Hertwig, R., & Todd, P. M. (2003). More is not always better: The benefits of cognitive limits. In D. Hardman and L. Macchi (eds.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making* (p. 213–231). Chichester, UK: Wiley.
- Hilbig, B. E. (2010). Reconsidering "evidence" for fast-and-frugal heuristics. *Psychonomic Bulletin & Review*, 17, 923–930. <http://dx.doi.org/10.3758/PBR.17.6.923>
- Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2010). One-reason decision making unveiled: A measurement model of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 123–134. <http://dx.doi.org/10.1037/a0017518>
- Hilbig, B. E., Scholl, S. G., & Pohl, R. F. (2010). Think or blink — is the recognition heuristic an "intuitive" strategy? *Judgment and Decision Making*, 5, 300–309.
- Hogarth, R. M., & Karelaia, N. (2005). Ignoring information in binary choice with continuous variables: When is less "more"? *Journal of Mathematical Psychology*, 49, 115–124. <http://dx.doi.org/10.1016/j.jmp.2005.01.001>
- Horn, S. S., Pachur, T., & Mata, R. (2015). How does aging affect recognition-based inference? A hierarchical Bayesian modeling approach. *Acta Psychologica*, 154, 77–85. <http://dx.doi.org/10.1016/j.actpsy.2014.11.001>
- Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G., & Gigerenzer, G. (2010). From recognition to decisions: Extending and testing recognition-based models for multialternative inference. *Psychonomic Bulletin & Review*, 17, 287–309. <http://dx.doi.org/10.3758/PBR.17.3.287>
- McCloy, R., Beaman, C. P., Frosch, C. A., & Goddard, K. (2010). Fast and frugal framing effects? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1043–1052. <http://dx.doi.org/10.1037/a0019693>
- Newell, B. R., & Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 923–935. <http://dx.doi.org/10.1037/0278-7393.30.4.923>
- Newell, B. R. (2011). Recognising the recognition heuristic for what it is (and what it's not). *Judgment and Decision Making*, 6, 409–412.
- Oeusoonthornwattana, O., & Shanks, D. R. (2010). I like what I know: Is recognition a non-compensatory determiner of consumer choice? *Judgment and Decision Making*, 5, 310–325.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349. <http://dx.doi.org/10.1126/science.aac4716>
- Oppenheimer, D. M. (2003). Not so fast! (and not so frugal!): Rethinking the recognition heuristic. *Cognition*, 90, B1–B9. [http://dx.doi.org/10.1016/S0010-0277\(03\)00141-0](http://dx.doi.org/10.1016/S0010-0277(03)00141-0)
- Pachur, T., Bröder, A., & Marewski, J. N. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making*, 21, 183–210. <http://dx.doi.org/10.1002/bdm.581>
- Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 983–1002. <http://dx.doi.org/10.1037/0278-7393.32.5.983>
- Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. G. (2011). The recognition heuristic: A review of theory and tests. *Frontiers in Psychology*, 2, 1–14. <http://dx.doi.org/10.3389/fpsyg.2011.00147>
- Pohl, R. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, 19, 251–271.
- Richter, T., & Späth, P. (2006). Recognition is used as one cue among others in judgment and decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 150–162. <http://dx.doi.org/10.1037/0278-7393.32.1.150>