

The limited use of the fluency heuristic: Converging evidence across different procedures

Rüdiger F. Pohl¹ · Edgar Erdfelder¹ · Martha Michalkiewicz^{1,2} · Marta Castela¹ · Benjamin E. Hilbig³

Published online: 6 July 2016
© Psychonomic Society, Inc. 2016

Abstract In paired comparisons based on which of two objects has the larger criterion value, decision makers could use the subjectively experienced difference in retrieval fluency of the objects as a cue. According to the fluency heuristic (FH) theory, decision makers use fluency—as indexed by recognition speed—as the only cue for pairs of recognized objects, and infer that the object retrieved more speedily has the larger criterion value (ignoring all other cues and information). Model-based analyses, however, have previously revealed that only a small portion of such inferences are indeed based on fluency alone. In the majority of cases, other information enters the decision process. However, due to the specific experimental procedures, the estimates of FH use are potentially biased: Some procedures may have led to an overestimated and others to an underestimated, or even to actually reduced, FH use. In the present article, we discuss and test the impacts of such procedural variations by reanalyzing 21 data sets. The results show noteworthy consistency across the procedural variations revealing low FH use. We discuss potential explanations and implications of this finding.

Keywords Judgment · Decision making · Retrieval cues · Fluency

✉ Rüdiger F. Pohl
pohl@psychologie.uni-mannheim.de

¹ Department of Psychology, School of Social Sciences, University of Mannheim, Schloss Ehrenhof Ost, D-68131 Mannheim, Germany

² Institute for Experimental Psychology, Heinrich-Heine-University, Düsseldorf, Germany

³ Cognitive Psychology, University of Koblenz-Landau, Koblenz, Germany

In typical tasks used to study inferential decision making, participants receive pairs of objects and are asked to infer which of the two objects has the larger criterion value—for example, which of two cities is more populous (e.g., Goldstein & Gigerenzer, 2002; Pohl, 2006). Given that participants typically do not possess any direct criterion knowledge, they need to infer the criterion value from further probabilistic knowledge (termed “cues”—for example, whether or not the cities have an international airport. In such a situation, participants could also rely on the fluency of memory retrieval (of an object’s name) as a cue, inferring that the object retrieved more speedily has the larger criterion value. This is what Schooler and Hertwig (2005) proposed as the “fluency heuristic” (FH) a decade ago. In the next section, we describe both their approach and the different methods to measure FH use. Then we discuss the procedures used in studies that have investigated the FH so far, and argue that, for methodological reasons, some of these studies may have overestimated FH use, whereas others may have underestimated it, or actually changed the FH use itself. Next, we test the impacts of such procedural variations by reanalyzing the data from several studies that were based on different procedures. The main result of this endeavor is that FH use is consistently low, despite some critical procedural variations. In the final section, we discuss potential explanations of our findings and more general implications for fluency as a cue in inferential decision making.

The fluency heuristic

Schooler and Hertwig (2005) introduced the FH as a simple, one-reason inferential strategy that could be applied whenever, in a pair of recognized objects, one of the object names is retrieved more speedily from memory than the other, thus

leading to a subjectively perceivable difference in retrieval fluency (see Herzog & Hertwig, 2013, for a recent overview). Schooler and Hertwig suggested that a minimum retrieval-time difference of 100 ms could be detected. Object pairs that conform to this condition are called *fluency-heterogeneous*. This notion reflects the finding that not the absolute levels of fluency, but rather *differences* in fluency, are often the more powerful cue (see, e.g., Wänke & Hansen, 2015; Whittlesea & Williams, 2000). In this case, the object retrieved more speedily should be chosen, on the basis of fluency alone and irrespective of any further cue knowledge or other information. As a proxy for retrieval fluency, Schooler and Hertwig proposed the recognition speed of each object, which can be assessed in a separate recognition test (see Hertwig, Herzog, Schooler, & Reimer, 2008). So far, all empirical studies investigating the FH have used recognition speed as a proxy for retrieval fluency.

In some studies, the adequacy of the FH as a description of decision-making behavior was assessed by simply counting how often the FH predicted participants' choices correctly—that is, how often the more speedily recognized object was indeed chosen. This *adherence* (or *accordance*) rate is typically substantial and reliably above chance level. For example, Hertwig et al. (2008, Exp. 3) reported values of .63, .68, and .74 (for different content domains). Similarly, in two of our own studies (Hilbig, Erdfelder, & Pohl, 2011, and Hilbig & Pohl, 2009, Exp. 3, as reanalyzed by Hilbig, 2010), the adherence rates were .63 and .68, respectively. That is, in about two thirds of all potential cases, participants' choices were as predicted by the FH. However, the adherence rate does not provide a suitable estimate of how often participants actually *used* the FH, because the more speedily recognized object could have been chosen for a number of reasons, only one of them being fluency differences alone. In other words, the adherence rate is a confounded measure, contaminated by the influences of several distinct processes (cf. Bröder & Schiffer, 2003; Fiedler, 2010; Goldstein & Gigerenzer, 2002; Hilbig, 2010; Hilbig & Pohl, 2008).

To remedy this problem and disentangle the different processes that lead to choosing the more speedily recognized object, Hilbig et al. (2011) proposed a multinomial processing-tree model, the *r-s model*. Models of this type generally explain observable categorical data (like choice frequencies) through a set of latent parameters (see Erdfelder et al., 2009, for an overview). In this case, the parameters represent the probabilities of underlying cognitive processes leading to the observed outcomes. Model fits and differences in the parameter values can be tested via maximum-likelihood techniques (see Hu & Batchelder, 1994, and Batchelder & Riefer, 1999, for details). The *r-s model* comprises several such parameters, of which the parameter *s* is the one that is primarily relevant for our present purpose. This parameter provides an uncontaminated measure of FH use—that is, the

probability of choosing the more speedily recognized object (in fluency-heterogeneous pairs) on the basis of fluency differences alone, and thus ignoring any other knowledge or information. Estimates of *s* are typically much lower than adherence rates. Hilbig et al. (2011) reported an average *s* estimate of only .23. Similarly, Schwikert and Curran (2014) found *s* estimates of .16 (Exp. 1) and .21 (Exp. 2). All of these values are rather low, but significantly larger than zero (as tested via the decrement in fit of the *r-s* model when assuming *s* = 0). Thus, we may conclude that the FH was indeed applied, but in a rather small portion of the potentially applicable cases only (see also Marewski & Schooler, 2011, for similar conclusions). However, except for Hertwig et al.'s (2008) Experiment 3, all of the studies designed (or later reanalyzed) to investigate FH use (i.e., Hertwig, Pachur, & Kurzenhäuser, 2005, Exp. 2; Hilbig & Pohl, 2009, Exp. 3; Hilbig et al., 2011; Marewski & Schooler, 2011, Exps. 1–3; Pachur & Hertwig, 2006, Exp. 2; Schwikert & Curran, 2014, Exps. 1 and 2; Volz, Schooler, & von Cramon, 2010) could be criticized for having adopted experimental procedures that may have led to biased estimates of FH use—namely, either to an overestimation of FH use, thus suggesting that the true probability of FH use is even lower, or to a reduction or underestimation of FH use, thus suggesting that the true probability of FH use is larger than those proposed by Hilbig et al. (2011) and Schwikert and Curran (2014). We discuss these problems in the next section.

Experimental procedures and their potential impacts

A typical study testing the FH (and other heuristics) involves a recognition task and a paired-comparison inference task. In the recognition task, participants receive a list of objects (e.g., cities) in random or alphabetical order and judge for each object whether or not they recognize it. The recognition times are measured and used as a proxy for retrieval fluency. In the paired-comparison task, participants receive a list of pairs, created using the same set of objects (e.g., cities). For each of these pairs, they are asked to infer which of the two objects has the larger criterion value (e.g., population).

A closer look at such studies reveals that two of the procedures may potentially bias measures of FH use, or may even affect FH use itself, namely (a) the task order, in several studies that have assessed retrieval fluency via a recognition task performed *after* the inference task (and not *prior* to it), and (b) the repeated presentation of objects, in several studies that have presented the same objects *repeatedly* during the inference task (rather than only *once*). In the following sections, we consider the potential problems induced by these procedures

in detail. Both may arguably have led to a biased assessment of FH use.

Problem 1: Task order

In several studies, the recognition task followed rather than preceded the inference task (Hertwig et al., 2008, Exp. 4; Hilbig & Pohl, 2009, Exp. 3; Hilbig et al., 2011; Marewski & Schooler, 2011, Exps. 1–3; Pachur & Hertwig, 2006, Exp. 2; Volz et al., 2010). One reason for choosing this task order could be to prevent alerting participants to specific cues of the object set (like recognition or fluency) prior to the inference task. However, placing the recognition task at the end necessarily means that every object has already been presented, and—if recognized—retrieved at least once from memory during the inference task—that is, before the recognition times are assessed. Presentation and subsequent retrieval both will arguably strengthen the memory representations of the respective objects. As a consequence, retrieval times would on average be shorter (i.e., objects would be recognized more fluently) if the recognition task is placed at the end rather than at the beginning (henceforth referred to as *Proposition 1*). If one further assumes that retrieval times follow the power law of practice (or an exponential decay function; cf. Heathcote, Brown, & Mewhort, 2000) as is standard in theories of repeated memory retrieval (see, e.g., Anderson & Schooler, 1991; Schooler & Hertwig, 2005), then the differences in retrieval times between recognized objects will, on average, be smaller if the recognition task is placed at the end (*Proposition 2*). Volz et al. (2010, p. 832) voiced this concern as follows: “We would expect that the recognition latencies would be faster, because the cities [that were used as materials] would have been recently seen in the inference task. As a result, we may be underestimating the absolute differences in retrieval fluency between items.” If, in turn, retrieval-time differences fall below the critical threshold, then the number of pairs identified as fluency-heterogeneous (i.e., pairs with sufficiently different retrieval times) will also be smaller if the recognition task is given at the end (*Proposition 3*), because an unknown number of objects may have approached their lower asymptotic retrieval-time level. In other words, the set of potential FH pairs is artificially reduced and does not capture all truly fluency-heterogeneous pairs in the preceding inference task. Correspondingly, and assuming that participants base their decisions on the experienced fluency differences in the inference task, the probability of FH use may well be overestimated relative to a condition in which the recognition task comes first (*Proposition 4*). This occurs simply because not all of the actually fluency-heterogeneous pairs in the inference task may be uncovered by the later recognition task. As a consequence, FH use would be estimated on only a subset of those pairs in which an individual might actually have used the FH, but this subset most likely consists of the

object pairs that originally showed the largest retrieval-time differences. Hilbig et al. (2011) reported that use of the FH increased with the retrieval-time difference, as does the validity of fluency as a cue (Hertwig et al., 2008). Hence, the probability of FH use may be overestimated if the subset of FH-applicable cases is determined *after* the inference task.

Problem 2: Repeated presentations

The second procedural variation mentioned above could lead to an underestimation of FH use, or even to a reduction in actual FH use, thus producing a result opposite to that from Proposition 4. Several studies have used exhaustive pairings of all objects in the inference task (Hilbig & Pohl, 2009, Exp. 3; Hilbig et al., 2011; Pachur & Hertwig, 2006, Exp. 2; Schwikert & Curran, 2014, Exps. 1 and 2). This means that, for example, if 25 objects are included, each object is presented 24 times throughout the inference task. As a consequence, the time to retrieve each object from memory (as the basis for subjectively experienced fluency) will decrease with repeated presentations (henceforth referred to as *Proposition 5*), and so will the differences in retrieval times between recognized objects (*Proposition 6*). The preconditions for these two propositions to hold are the same as for Problem 1. Again, assuming that more and more fluency differences fall below the critical threshold, the number of pairs to which the FH could be applied (i.e., truly fluency-heterogeneous pairs) will continuously decrease during the inference task (*Proposition 7*). Note that, in contrast to Problem 1, in which the size of the diagnosed set of fluency-heterogeneous pairs differed depending on the task order, the size of this set is constant for the whole inference task, because recognition times are assessed only once. Thus, if the subjectively perceived set of fluency-heterogeneous pairs actually diminishes substantially with repeated presentations (as Proposition 7 assumes), but the identified set of FH pairs remains constant (which is the case), the relative frequency of FH use must decline—that is, FH use is underestimated (*Proposition 8*). Still another possibility is that FH use actually declines due to the repeated object presentations: With a reduced number of fluency-heterogeneous pairs, the discrimination rate of fluency (i.e., the proportion of pairs in which fluency distinguishes between alternatives) also drops. In turn, the success rate of fluency (i.e., the proportion of correct decisions when using fluency as the only cue throughout the task) also declines, when seen across the whole set of pairs (Martignon & Hoffrage, 1999; Newell, Rakow, Weston, & Shanks, 2004). In sum, if truly fluency-heterogeneous pairs decline across the task, the effort of strategy switching increases, which may motivate decision makers to discard the FH and rely on some other strategy that is more often applicable, and thus entails fewer switching costs (see, e.g., Bröder & Schiffer, 2006). As a consequence, decision makers might not engage in FH use anymore.

An additional danger associated with repeated presentations of the same objects was pointed out by Schweickart and Brown (2014, p. 285; see also Pohl, 2011):

Because the same items are repeatedly presented across different pairs, it is possible that, during the course of the experiment, people create ad hoc cognitive structures that represent the linear ordering of the items used in the paired comparison task. As a result, participants might rely predominantly on these “temporary data sets” in their comparative judgments, instead of retrieving information from semantic memory anew on each trial.

That such linear orders are indeed powerful structures in inferential decision making was shown by Brown and Tan (2011), Pohl and Hilbig (2012), and Schweickart and Brown (2014). Subjective retrieval fluency might be helpful to construct such an order in the first place, but not to infer the answers in single pairs, at least not after such an order has been established—which, in turn, is arguably fostered by repeating the objects. As a consequence, use of the FH may decrease with repeated presentations of the same objects, which would thus contribute to a generally low estimate of FH use.

Empirical evidence

A few studies have reported results that are relevant to the two problems discussed above. Hertwig et al. (2008, Exp. 3) varied the task order and assessed the accordance rates of both the recognition heuristic and the FH. They found that “the order of the recognition and inference tasks had no statistically significant effect on the accordance to the fluency and recognition heuristics. . . . None of the implications of the reported analysis changed when analyzing the two task orders separately” (p. 1199). Similarly, Schwikert and Curran (2014) also varied the task order in their Experiment 1 while assessing FH use via the $r-s$ model (Hilbig et al., 2011). They found, for two different materials (U.S. cities and world countries), that the probability of using the FH did not depend on whether the recognition task came first or last. So the effect of task order seems negligible, at least in these two studies.

The latter study (Schwikert & Curran, 2014) is also the only one so far that has addressed the problem of repeated presentations of objects. The authors presented each object exactly four times, separated in four blocks of trials. They then analyzed FH use for the first and last blocks of trials (i.e., for the first and fourth presentations of each object) and found estimates of .15 and .14, respectively, in Experiment 1, and .20 and .20 in Experiment 2. As such, repetition did not appear to have any noteworthy influence on FH use.

In sum, the few available data suggest that neither the task order nor the repeated presentation of objects in the inference task has an impact on FH use or on estimates thereof. However, the evidence so far is too scarce, is not always suited to reaching a firm conclusion, and only one study has addressed both questions (Schwikert & Curran, 2014). Thus, we set out to reanalyze a large set of studies to replicate the reported findings, preferably across a large set of experiments, and thus to critically test whether the described experimental procedures may have led to biased assessments of FH use. We identified 21 recent studies, mainly from our own lab, that appeared suitable because they had manipulated the crucial experimental procedures discussed above, thus allowing for within-experiment comparisons. In this way, we extend the findings from Schwikert and Curran by including more diverse experimental situations—namely, by varying the number of repetitions in the inference task (from two to 24), by varying the number of trials preceding the final recognition test (from 84 to 300), and by varying the type of material (see Table 1). Thus, our conclusions are less limited to a specific experimental setup. Information on the studies’ sample sizes, the materials used, and the decision-making criteria are provided in Table 1.

Problem 1: Are there effects of task order?

Of the 21 studies listed in Table 1, 17 included one condition with the recognition task given before the inference task and another condition with the recognition task given after the inference task. We compared these two conditions on a number of measures (see Tables 2, 3 and 4) following the four propositions outlined above.

Proposition 1 (decrease in recognition times) To assess recognition times (as the standard proxy for retrieval times), we computed the median recognition times per participant in both task orders. The means of these medians and statistical tests of their difference (one-tailed t tests) are given in Table 2. The results provide a clear picture: In 15 of the 17 data sets, recognition times were shorter (on average across all studies by 219 ms) whenever the recognition task followed rather than preceded the inference task. This difference was statistically significant in 12 of the 15 studies. Thus, the data clearly corroborate Proposition 1, that recognition times decrease due to repeated retrieval.

Proposition 2 (decrease in recognition-time differences) Next we computed the mean recognition-time difference per participant for all pairs with both objects recognized. The overall means per task order and a statistical

Table 1 Sources and description of the 21 reanalyzed data sets

Data Set	Sample Size	Materials; Number of Pairs; and Decision Criterion	# of Data ^a
Castela & Erdfelder (in press)			
1	Exp. 1 39	80 of the 150 most populous U.S. cities; 240 pairs; population size?	9,360
2	Exp. 2 33	80 of the 150 most populous U.S. cities; 240 pairs; population size?	7,920
Hilbig, Erdfelder, & Pohl (2011)			
3	Exp. 1 66	14 most populous cities of Poland and Austria (excl. the largest ones); 182 pairs; population size?	12,012
Hilbig, Michalkiewicz, Castela, Pohl, & Erdfelder (2015)			
4	Exp. 1 23	20 out of the 61 most populous world cities (control group); 190 pairs; population size?	4,370
5	Exp. 2 22	20 out of the 61 most populous world cities (control group); 190 pairs; population size?	4,180
6	Exp. 3 32	84 out of the 100 most populous world cities (control group); 84 pairs; population size?	2,688
Michalkiewicz & Erdfelder (2016)			
7	Exp. 1 64	100 out of the 150 most populous U.S. cities (first session); 300 pairs; population size?	19,200
8	Exp. 2 83	25 out of the 61 most populous world cities (first session); 300 pairs; population size?	24,900
9	Exp. 3a 68	25 out of the 100 most successful celebrities 2012; 300 pairs; success?	20,400
10	Exp. 3b 68	25 out of the 100 most successful German movies; 300 pairs; success?	20,400
11	Exp. 3c 64	25 out of the 60 largest islands worldwide; 300 pairs; size?	19,200
12	Exp. 3d 64	25 out of the 100 most successful musicians; 300 pairs; success?	19,200
13	Exp. 4a 87	25 out of the names of the 100 most successful celebrities 2012; 300 pairs; success?	26,100
14	Exp. 4b 87	25 out of the pictures of the 100 most successful celebrities 2012; 300 pairs; success?	26,100
Michalkiewicz, Arden, & Erdfelder (2016)			
15	Exp. 1a 44	25 out of the 100 most successful celebrities 2012; 300 pairs; success?	13,200
16	Exp. 1b 48	25 out of the 100 most successful celebrities 2012; 300 pairs; age?	14,400
Michalkiewicz, Minich, & Erdfelder (2016)			
17	Exp. 1a 80	25 out of the 60 longest rivers; 300 pairs; length?	24,000
18	Exp. 1b 80	25 out of the 100 most successful musicians; 300 pairs; success?	24,000
Schwikert & Curran (2014) ^b			
19	Exp. 1a 24	100 most populous U.S. cities; 200 pairs; population size?	4,800
20	Exp. 1b 24	100 most populous world countries; 200 pairs; population size?	4,800
21	Exp. 2 34	100 most populous U.S. cities and 100 most populous world countries; 200 pairs each; population size?	13,600

^a The number of data is given by the product of the sample size and number of pairs. ^b We thank Shane Schwikert and Tim Curran for providing us with the raw data from their two experiments.

test of their difference (one-tailed t test) are also given in Table 2. The results show the predicted trend: In 12 of the 17 studies, the mean recognition-time difference between the recognized objects was smaller when the recognition task followed rather than when it preceded the inference task. The mean overall difference decreased on average (across all studies) by 69 ms, from 404 to 335 ms. Out of the 12 studies that showed such a decrease, seven yielded a statistically significant difference. Thus, the data are compatible with Proposition 2. However, note that the mean difference in recognition times was still substantially above the critical limit of 100 ms in all conditions.

Proposition 3 (decrease in number of fluency-heterogeneous pairs) We computed the mean proportions of fluency-heterogeneous pairs (i.e., with a recognition-time difference above 100 ms) out of all pairs with both

objects recognized for each task order. These values correspond to parameter $1-p$ in the $r-s$ model of Hilbig et al. (2011), and are given in Table 3. We then duplicated the $r-s$ model to capture both task orders (with two p estimates, one for each task order). To test whether the two p parameters differed significantly, we set them equal to each other and tested the resulting decrement in model fit (ΔG^2). The results are again clear: The proportion of fluency-heterogeneous pairs was substantially smaller whenever the recognition task followed rather than preceded the inference task. Out of the 17 included studies, 13 showed this result, and all were statistically significant. Thus, the data clearly speak for Proposition 3. In absolute terms, however, the decrease appeared to be rather small: On average, the proportion dropped by only .04, from .80 to .76. Thus, the large majority of pairs of recognized objects remained classified as fluency-heterogeneous, even if the recognition task was placed at the end.

Table 2 Recognition times and recognition-time differences depending on the task order (Problem 1, Propositions 1 and 2) for the 17 included data sets: Means of the median recognition times (in milliseconds) for both task orders, real difference, and test of the difference (one-tailed *t* test)

Data Set	Recognition Times (ms)				Recognition-Time Differences (ms)			
	Recognition Task First	Recognition Task Last	Difference	<i>t</i> Test <i>df</i> , <i>t</i> , <i>p</i> , <i>d</i>	Recognition Task First	Recognition Task Last	Difference	<i>t</i> Test <i>df</i> , <i>t</i> , <i>p</i> , <i>d</i>
4	1,044	967	-77	21, 0.81, .21, 0.34	240	335	95	21, 1.36, .91, 0.57
5	902	1,052	150	20, 1.61, .94, 0.69	255	449	194	20, 2.20, .98, 0.94
6	833	849	16	30, 0.36, .64, 0.13	207	281	74	30, 2.13, .98, 0.75
7	804	702	-102	62, 3.71, <.001*, .94	217	180	-37	62, 1.57, .06, 0.40
8	1,224	902	-322	81, 4.99, <.001*, 1.09	492	314	-178	81, 3.48, <.001*, 0.76
9	1,162	921	-241	66, 2.98, .002*, 0.72	432	300	-132	66, 2.41, .009*, 0.59
10	1,366	997	-369	66, 4.41, <.001*, 1.07	529	357	-172	66, 2.96, .002*, 0.72
11	1,344	931	-413	62, 5.45, <.001*, 1.36	627	365	-262	62, 3.80, <.001*, 0.95
12	1,116	917	-199	62, 3.11, .001*, 0.78	397	318	-79	62, 1.38, .09, 0.35
13	1,252	893	-359	85, 6.09, <.001*, 1.31	532	377	-155	85, 2.77, .003*, 0.59
14	1,260	908	-352	85, 4.99, <.001*, 1.07	626	355	-271	85, 3.60, <.001*, 0.77
15	1,203	979	-224	42, 2.76, .004*, 0.84	464	531	67	42, 0.51, .69, 0.15
16	1,324	949	-375	46, 3.71, <.001*, 1.07	500	353	-147	46, 1.74, .045*, 0.50
17	1,087	793	-294	78, 5.44, <.001*, 1.22	443	356	-87	78, 1.03, .15, 0.23
18	1,154	926	-228	78, 4.28, <.001*, 0.96	421	363	-58	78, 1.28, .10, 0.29
19	1,034	922	-112	46, 0.79, .22, 0.23	240	249	9	46, 0.18, .57, 0.05
20	1,029	817	-212	46, 1.39, .09, 0.39	243	208	-35	46, 0.91, .19, 0.26

* Significant difference ($p < .05$). The effect-size measure is Cohen's *d*.

Table 3 Fluency-heterogeneous pairs and probability of FH use, depending on the task order (Problem 1, Propositions 3 and 4), for the included 17 data sets: Model fits, estimates of the proportions of fluency-heterogeneous pairs (parameter $1-p$) for both task orders, the difference

test); and mean recognition-time differences for pairs of recognized objects (in milliseconds) for both task orders, the real difference, and test of the difference (one-tailed *t* test)

Data Set	Model Fit $G^2(df=2)$, <i>p</i>	Proportions of Fluency-Heterogeneous Pairs				Probabilities of FH Use			
		Recognition Task First	Recognition Task Last	Difference Δp	Model Test $\Delta G^2(df=1)$, <i>p</i>	Recognition Task First	Recognition Task Last	Difference Δs	Model Test $\Delta G^2(df=1)$, <i>p</i>
4	0.78, .68	.74	.78	.04	2.56, .11	.05	.29	.24	11.43, <.001*
5	0.11, .94	.74	.80	.06	7.87, .005*	.09	.24	.13	7.92, .004*
6	5.33, .07	.70	.76	.06	4.94, .03*	.16	.13	-.03	0.19, .66
7	1.32, .52	.70	.62	-.08	33.70, <.001*	.30	.30	.00	0.02, .88
8	1.57, .46	.83	.78	-.05	45.40, <.001*	.20	.26	.06	8.30, .004*
9	8.11, .02†	.81	.78	-.03	8.87, <.001*	.22	.21	-.01	0.13, .72
10	4.21, .12	.84	.80	-.04	20.73, <.001*	.16	.25	.09	15.71, <.001*
11	20.41, <.001†	.87	.79	-.08	69.89, <.001*	.27	.27	.00	0.01, .92
12	5.36, .07	.82	.75	-.07	58.00, <.001*	.29	.28	-.01	0.08, .78
13	4.12, .13	.86	.80	-.06	45.71, <.001*	.21	.22	.01	0.29, .59
14	22.50, <.001†	.85	.80	-.05	20.71, <.001*	.22	.29	.07	6.85, <.001*
15	2.66, .27	.81	.81	.00	0.45, .50	.25	.22	-.03	0.67, .41
16	97.61, <.001†	.85	.77	-.08	39.81, <.001*	.00	.00	.00	0.00, 1.00
17	16.44, <.001†	.83	.80	-.03	102.70, <.001*	.20	.20	.00	0.08, .78
18	13.46, .001†	.84	.73	-.11	20.18, <.001*	.30	.27	-.03	2.22, .14
19	1.37, .50	.74	.68	-.06	27.23, <.001*	.15	.18	.03	1.91, .17
20	11.57, .003†	.72	.68	-.04	9.04, .002*	.16	.14	-.02	0.54, .46

† No satisfactory model fit ($p < .05$). * Significant difference ($p < .05$).

Δp (as an effect-size measure; Moshagen, 2010), and test of the difference; and estimates of FH use (parameter *s*) for both task orders, the difference Δs (as an effect-size measure; Moshagen, 2010), and test of the difference

Table 4 FH use depending on object repetitions (Problem 2, Proposition 8) for the 21 included data sets: Number of object presentations (k), model fit, estimates of FH use (parameter s) for the

first and last presentations of objects, the difference Δs (as an effect-size measure; Moshagen, 2010) and test of the difference

Data Set	k	Model Fit $G^2(df=2), p$	Probability of FH Use			
			First Object Presentation	Last Object Presentation	Difference Δs	Model Test $\Delta G^2(df=1), p$
1	6	2.60, .27	.26	.20	-.06	0.67, .41
2	6	0.38, .83	.36	.29	-.07	0.82, .37
3	13	1.50, .22	.27	.15	-.12	1.73, .19
4	19	2.12, .34	.33	.00	-.33	3.98, .05*
5	19	0.41, .81	.33	.00	-.33	4.17, .04*
6	2	1.83, .40	.14	.15	.01	0.05, .83
7	6	1.18, .55	.31	.32	.01	0.02, .90
8	24	0.64, .73	.22	.29	.07	1.01, .31
9	24	1.02, .60	.31	.13	-.18	2.87, .09
10	24	0.02, .99	.19	.17	-.02	0.02, .88
11	24	1.32, .52	.18	.36	.18	3.44, .06
12	24	1.33, .51	.35	.39	.04	0.22, .63
13	24	1.27, .53	.23	.25	.02	0.02, .88
14	24	7.38, .02†	.17	.18	.01	0.02, .91
15	24	0.98, .61	.18	.21	.03	0.06, .81
16	24	12.37, <.01†	.00	.00	.00	0.00, 1.00
17	24	5.48, .06	.13	.23	.10	0.84, .36
18	24	1.45, .48	.27	.38	.11	2.86, .09
19	4	1.78, .41	.16	.15	-.01	0.09, .76
20	4	2.47, .29	.14	.13	-.01	0.10, .75
21	4	3.14, .08	.20	.20	.00	0.98, .32

* No satisfactory model fit ($p < .05$). † Significant difference ($p < .05$).

Proposition 4 (overestimation of FH use) The probabilities of FH use were estimated via parameter s of the $r-s$ model, again including both task orders. These estimates are summarized in Table 3. The difference in FH use between the two task orders was tested by setting the two s parameters in the $r-s$ model equal to each other and testing the resulting decrement in model fit (ΔG^2). The results showed that estimates of FH use did not differ depending on whether the recognition task was placed at the beginning or at the end. The corresponding average probabilities were .19 and .22, respectively, with 12 of the 17 studies showing no significant difference (and also no consistent numerical trend). Only five of the 17 data sets were in line with Proposition 4; that is, they showed a significantly larger probability of FH use when the recognition task was given *after* the inference task. Thus, there is only very limited evidence for Proposition 4, and we are left to conclude that it does not hold. Note that this conclusion would not change when discarding data sets that yielded model misfits.

In sum, we found that placing the recognition test at the end rather than the beginning of the procedure indeed produces

shorter recognition times, thus decreasing recognition-time differences and the number of fluency-heterogeneous pairs in the decision phase of the experiments. However, it does not affect the estimated probabilities of FH use. Uncontaminated measures such as the s parameter of the $r-s$ model suggest that FH use is consistently rare, irrespective of the position of the recognition task in the task sequence.

Problem 2: Are there effects of repeated object presentations?

The objects in all 21 data sets listed in Table 1 were presented repeatedly. However, testing Propositions 5–7 turned out to be a challenge, and only Proposition 8 could be tested in a straightforward way.

Propositions 5–7 (decreases in recognition times, recognition-time differences, and numbers of fluency-heterogeneous pairs) Unfortunately, we do not have any direct measures of retrieval fluency for pairs of objects, let alone differences between objects or changes in retrieval fluency.

The only data that have been assessed during the inference task are decision times and choices. Decision times, however, include the retrieval times for *both* objects (plus other processes), so that not much can be derived from them regarding the retrieval fluency for each of the two objects (cf. Marewski & Mehlhorn, 2011; Marewski & Schooler, 2011). Thus, we lack conclusive data to test Proposition 5, and as a consequence, also Propositions 6 and 7.

Nevertheless, for several reasons, the case can be made that all three propositions are likely to hold. First, the data analysis for the task order (see Problem 1 above) confirmed that the repeated presentation of objects (i.e., following the inference task) led to shorter recognition times, smaller recognition-time differences, and fewer fluency-heterogeneous pairs. Thus, by analogy, it seems plausible that the same changes would hold due to repeated presentations *during* the inference task. Second, we found two indirect measures that provide some evidence: (1) We computed the mean decision times for fluency-heterogeneous pairs in the first and last bins of trials (with each consecutive bin containing the next presentation of an object; see below). Corresponding data were available for 17 of the 21 data sets. Decision times declined significantly in all 17 studies, on average by 952 ms, from the first to the last bin of trials. Also, (2) one recent study (Castela & Erdfelder, *in press*) repeated the recognition task for the same objects three times (in separate sessions). We reanalyzed the data and found significantly decreasing recognition times in each of two experiments. In sum, both of these findings—decreasing decision times and decreasing recognition times—are in line with the predictions of diminishing retrieval times (Proposition 5) and the subsequent consequences (as stated in Propositions 6 and 7). Note, however, that these findings cannot be taken as direct evidence, since the observed declines may also have been due to other processes (like practice effects or reduced motivation). Further research would be needed to evaluate Propositions 5–7 more directly. Nonetheless, the most central prediction (Proposition 8) can be tested conclusively.

Proposition 8 (underestimation of FH use) For each of the 21 studies, we split the set of inference trials into consecutive bins according to the number of object presentations. Whereas Schwikert and Curran (2014) used and repeated a set of pairs that contained exactly one presentation of each object, the other studies used random lists of all possible pairs, such that each bin contained approximately one presentation of each object. The number of presentations of each object varied from 2 to 24. We estimated the probability of FH use with the *r-s* model (Hilbig et al., 2011), separately for the first and the last bin (see Table 4). As before, differences in FH use were assessed by setting the two *s* parameters for the first and last bins equal to each other and testing the resulting decrement in model fit (ΔG^2).

The results were again clear-cut in showing no difference in FH use for the first versus the last bin. The mean estimates of FH use were .23 and .20, respectively. Of the 21 studies, only two were in line with Proposition 8, showing a statistically significant decrease in FH use from the first to the last bin, whereas the remaining 19 showed no significant difference (and also no consistent numerical trend). Again, removing the two nonfitting data sets left the results unaltered. Thus, we conclude that Proposition 8 is not valid. In a nutshell, FH use proved to be invariant against repeated object presentations in the binary decision task.

Discussion

In this article, we set out to test whether two typical experimental procedures (recognition task *after* inference task and repeated presentation of objects) may have led to a biased estimation of FH use. To this end, we reanalyzed the data sets of 21 studies that allowed us to assess the impact of task order and repeated presentation of objects, thus aiming to replicate and extend earlier findings (Schwikert & Curran, 2014). By including more diverse experiments with respect to the number of object repetitions, number of trials preceding the final recognition test, and type of materials, our conclusions are less limited to a specific setup.

With respect to Problem 1, namely the order of tasks (i.e., recognition task before or after the inference task), we found that the mean recognition times were shorter (and thus fluency larger), the mean difference in recognition times for pairs of both objects recognized was smaller, and the proportion of fluency-heterogeneous pairs (with a minimum recognition-time difference of 100 ms) was smaller, when the recognition task followed rather than preceded the inference task. Thus, Propositions 1–3 were confirmed. However, by contrast, Proposition 4 was refuted: FH use was independent of task order, with 12 of the 17 data sets showing no difference in the probabilities of FH use across the two task orders. We thus consider the practice of assessing recognition times *after* the inference task just as appropriate as assessing them *before* (cf. Hertwig et al., 2008). In conclusion, the estimates of FH use reported previously can be considered valid, irrespective of when recognition times were assessed.

Similarly, for Problem 2, the repeated presentation (and, potentially, retrieval) of objects had little, if any, effect on estimates of FH use or on FH use itself. We compared FH use in the first and last presentations of objects in each of the 21 data sets, and found that 19 data sets did not show significant differences (nor even a consistent numerical difference), despite the fact that many of the studies presented objects as often as 24 times. Thus, Proposition 8 was clearly disconfirmed. In other words, whether objects are presented only once or several times during the inference task did not

appear to influence participants' FH use (or estimates thereof) systematically.

In conclusion, the effects of both of the identified procedural problems when assessing FH use seem practically negligible, so the typically small proportions of FH-based decisions, of only about 20 %, cannot be attributed to these experimental procedures. They should thus be considered unbiased. Nevertheless, open questions remain: namely, why the impact of these experimental procedures is so small, and more importantly, why FH use appears to be rare in general.

Small impact of experimental procedures

Arguably, the preexperimental familiarity differences between choice objects are so large and stable that both placing the recognition task at the end of the procedure and presenting objects repeatedly have only negligible impacts. For example, a German university student may have heard of San Francisco a few thousand times, but of San Antonio only a few hundred times (cf. Schooler & Hertwig, 2005). Thus, a comparably small number of repeated presentations of these cities during an experiment will not suffice to substantially increase their activation strengths in memory, and thus decrease retrieval times to such a degree that their actual differences are masked (cf. Volz et al., 2010).

Indeed, our analysis showed that detectable differences in fluency, and thus a large portion of fluency-heterogeneous pairs, remained (cf. Marewski & Schooler, 2011; Volz et al., 2010). On average, between 75 % and 80 % of all pairs when both objects were recognized were fluency-heterogeneous, and thus potential FH candidates. The observed decline of this proportion due to the task order (Problem 1) amounted to only 4 %. The decline of FH pairs due to the repeated presentations of objects (Problem 2) could hardly exceed this value, because all repetitions (from two to 24) occurred *before* the final recognition task. Thus, the decrease in the number of fluency-heterogeneous pairs was only minimal, so that participants had little reason to change their strategies.

Another, but at this point purely speculative, explanation for the small influence of procedural variations on FH use borrows from research on perceptual fluency. In this field, it is common to distinguish between “objective” and “subjective” fluency (Reber, Wurtz, & Zimmermann, 2004). Objective fluency is defined as some measure of processing speed, like reaction times. In contrast, subjective fluency, and thus the basis of subsequent behavior, may not only depend on the perceived ease of processing (as captured by objective measures), but also on additional factors, like the expected fluency, metacognitions, and attributional processes (see, e.g., Susser, Jin, & Mulligan, 2016, on metamnemonic beliefs). For example, participants may well differentiate whether experienced fluency stems from the retrieved object itself, or rather from some external source, the so-called *context*

(previous encounters, perceptual conditions, etc.; see Dechêne, Stahl, Hansen, & Wänke, 2009, 2010; Hansen & Wänke, 2013; Undorf & Erdfelder, 2015). Several studies have shown that such subjective evaluations of the experienced retrieval fluency could have substantial impact (see von Helversen, Gendolla, Winkielman, & Schmidt, 2008; Wänke & Hansen, 2015).

It is, however, questionable whether the influences that have been found for manipulations of perceptual fluency also apply to the procedural manipulations discussed here. One would have to assume that participants are aware that subjectively experienced increases in retrieval fluency during an experimental session are merely due to repeated presentations of that object, and are thus not diagnostic of one's true familiarity with that object (cf. Dechêne et al., 2009, 2010; Hansen & Wänke, 2013; Undorf & Erdfelder, 2015). Studying the effect of experimentally manipulated fluency on FH use, Hertwig et al. (2008, Exp. 4) found that adherence to the FH increased for more fluently processed items, but the effect was moderated by participants' memory of the fluency manipulation: Participants who remembered better which items had been manipulated showed less increase in adherence rates. This could be taken as evidence that people are indeed able to discount fluency as a cue. Such attributional processes, if shown to be real, could explain why subjective evaluations of fluency and its use as a cue may remain fairly constant, even if objective retrieval times decrease. Of course, so far this is merely a conjecture that will require further research.

Low utility (and use) of the FH

The core problem of the FH concept as introduced by Schooler and Hertwig (2005) appears to be that it limits FH use to very specific pairs of objects. First, both objects need to be judged as recognized, since the FH does not apply to pairs in which only one object is recognized or both objects are unrecognized. Second, the difference in retrieval times of the two objects has to be large enough to be subjectively detectable. Schooler and Hertwig (2005) pretested and then defined a minimum recognition-time difference of 100 ms. In pairs of recognized objects with smaller recognition-time differences, this difference would not be detected, and the FH could not be applied. Hilbig et al. (2011) also tested larger recognition-time differences of up to 1,000 ms as thresholds, but nonetheless found that FH use was rare, with a maximum estimate of .31 (see also Hertwig et al., 2008, Exp. 3).

A third limiting factor is that the validity of fluency as a cue is often comparatively low—that is, barely above chance (.50). The mean fluency validity was only .61 in the simulations by Schooler and Hertwig (2005), and ranged from .58 to .66 for five different domains in Study 1 of Hertwig et al. (2008). Furthermore, Volz et al. (2010) found a validity of

only .55 in their study. In a research overview (including 25 data sets), Herzog and Hertwig (2013) reported a mean fluency validity of .62, which is significantly above chance, but still relatively low. Similarly, the fluency validities were .57 and .60 in the recent Experiments 1 and 2 of Schwikert and Curran (2014). In our remaining 18 data sets, the fluency validity was on average .57 (ranging from .44 to .67).

Fourth, Marewski and Schooler (2011) suggested that the FH is more likely to be applied whenever two objects are merely recognized—that is, when no further knowledge can be retrieved. In all other cases, “when further knowledge about the objects is available, people seem to use knowledge-based strategies, which tend to be more accurate than the fluency heuristic for such cases” (Herzog & Hertwig, 2013, p. 204). The reason is that knowledge validity tends to be greater than fluency validity. For example, the validity of the knowledge cues that Gigerenzer and Goldstein (1996) identified for German cities mostly ranged from .71 to .91. In our 21 data sets, the knowledge validity for fluency-heterogeneous pairs of recognized cities was somewhat lower—namely, on average .65 (ranging from .57 to .74)—but was still greater than the fluency validity in 16 of the 21 data sets.

A fifth reason why evidence for FH use should be expected to be weak can be derived from the memory-state heuristic theory (Erdfelder, Küpper-Tetzl, & Mattern, 2011). According to this theory, three different memory states can underlie recognition judgments: (1) recognition certainty, (2) uncertainty (leading to guessing), and (3) rejection certainty. For decision criteria strongly related to object familiarity, it predicts that the object in the “higher” memory state would tend to be chosen, which should only hold more so, the larger the discrepancy in memory states (cf. Castela & Erdfelder, *in press*; Castela, Kellen, Erdfelder, & Hilbig, 2014). Since the FH applies to pairs of recognized objects only, the respective objects are often in the same memory state, and in some cases in adjacent states (i.e., recognition certainty for one object and uncertainty followed by guessing for the other). Hence, the maximum possible discrepancy in memory states is 1, whereas for other decisions strategies (like the recognition heuristic) it is 2 (i.e., recognition certainty and rejection certainty). By implication, if people follow the memory-state heuristic, preference for one of the objects to which the FH applies can only be weak.

In sum, all of these conditions (i.e., both objects need to be recognized, the subjectively experienced fluency difference must be sufficiently large, the fluency validity should be high, no further knowledge should be available, and discrepancy in the underlying memory states should be high) limit the utility of retrieval fluency as a single cue, which in turn may explain the observed low rates of FH use. This corresponds to the

conclusions of Marewski and Schooler (2011) and Herzog and Hertwig (2013), based on elaborate simulation studies within an ACT-R framework.

Marewski and Schooler (2011) proposed that “cognitive niches” (i.e., a limited number of optimal situations) exist for every decision-making strategy. The problem for applying the FH, according to Marewski and Schooler (p. 407), is that “the magnitude of recognition time differences correlates with the availability of knowledge, and as such, with the applicability of knowledge-based strategies.” In other words, in cases in which the FH could, in principle, be applied (due to large fluency differences), knowledge differences are likely to exist as well. Moreover, the knowledge differences are likely to be preferred as a decision basis, since knowledge validity tends to be larger than fluency validity (see above). In addition, when further knowledge is available for *both* objects in a pair, retrieval-speed differences tend to be small and hard to detect (Marewski & Schooler, 2011). As a consequence, the remaining “niche” for applying the FH consists solely of pairs of recognized objects with a large-enough retrieval-time difference, but without further knowledge. These cases, however, represent only a minority of all cases. Averaged across 16 published data sets (reported in Castela et al., 2014), only 21 % of all pairs in which both objects were recognized consisted of pairs in which both objects were “merely recognized”—that is, without having further knowledge (as indicated by participants’ judgments; see Pohl, 2006). Assuming that the overall proportion of fluency-heterogeneous pairs of 80 % (as averaged from the data in Table 3) also applies to the pairs of merely recognized objects, we would predict FH use of .17 (= .21 × .80), which corresponds fairly well to the typically estimated rates of FH use of around .20.

Assessment of fluency

So far, all research on the FH, including our own, has used recognition speed as a proxy for retrieval fluency. However, as we already outlined above, subjectively experienced fluency may well differ from this measure. Thus, a limiting factor for all of these studies could be that recognition speed is not a good proxy. In addition to the metamnemonic and attributional processes discussed above (see also Susser et al., 2016), the process of recognition may be composed of several subprocesses, some or all of which contribute more or less to subjectively experienced fluency (Reber et al., 2004). Thus, there are several options for how to operationalize objective fluency: For example, Benjamin, Bjork, and Schwartz (1998) used answer retrieval latencies; Reber et al. (2004), identification latencies; Schooler and Hertwig (2005), recognition latencies; Mueller, Dunlosky, Tauber, and Rhodes (2014), lexical

decision latencies; Susser et al. (2016), naming latencies; and Undorf and Erdfelder (2015), the numbers of trials to acquisition and self-paced study times (see also Koriat, 2008). This diversity of measures led Reber et al. (2004, p. 50) to claim that “there seemingly is no single objective fluency.” It even appears that different measures of fluency may lead to different results (Undorf & Erdfelder, 2015), because “no specific measure captures all aspects of processing fluency” (p. 655). Poldrack and Logan (1997) had already found that speed measures explained only a portion of subjective fluency in recognition.

Reber et al. (2004) drew a pessimistic conclusion, stating that “it remains unclear how objective speed is related to subjective experiences of fluency” (p. 47). In their own studies, they found that two subprocesses were only jointly related to subjective fluency (as measured by ratings of ease), but neither one alone. In other words, if researchers tap the wrong subprocess (or too few), not much can be learned about subjective fluency. Correspondingly, a large body of research has shown that subjectively experienced fluency might differ substantially from objective measures (see Greifeneder, Bless, & Pham, 2010; Hilbig, 2012; Lloyd, Westerman, & Miller, 2003; Newell & Shanks, 2007; Sanchez & Jaeger, 2015; Scholl, Greifeneder, & Bless, 2014; Unkelbach & Greifeneder, 2013; Westerman, Miller, & Lloyd, 2003; Whittlesea & Leboe, 2003).

In sum, these considerations underscore the necessity to better understand what “subjective fluency” actually is and how to assess it better (cf. Alter & Oppenheimer, 2009; Greifeneder et al., 2010; Hansen & Wänke, 2013; Herzog & Hertwig, 2013; Schwarz, 2004; Schwikert & Curran, 2014; Unkelbach & Greifeneder, 2013; Wänke & Hansen, 2015). Nonetheless, in all studies on the FH published so far, recognition speed has been taken as a proxy for fluency, so using new proxies would render our findings difficult to compare to those from previous studies. Besides, we would argue that recognition speed might still be a good proxy, at least in this domain. We did not manipulate perceptual fluency, so that neither object detection, readability, word/nonword decisions, or related processes played a role. We also did not manipulate conceptual fluency (by semantic priming or the like), so that recognition speed is a plausible proxy for the overall fluency of the underlying retrieval experience.

Recent research by Sanchez and Jaeger (2015) additionally showed that perceived fluency (as measured by subjective ratings of reading difficulty) did not relate to fluency-manipulation-based effects, whereas reading times (as an objective measure) did. Moreover, subjective and objective measures of fluency were uncorrelated. These findings question the role of subjective fluency and emphasize the role of objective measures. Still, we should bear in mind that recognition speed is only a

proxy, and possibly not the best one. With other measures of fluency, estimates of FH use might change.

Conclusions

The “niche” for potential application of the FH appears rather small, and thus the utility of fluency as the only cue rather limited. Most likely, only pairs in which both objects are “merely” recognized—that is, in which no further knowledge about them is available—are candidates for FH use. Therefore, the result that decision makers use the FH in only about 20 % of all pairs in which both objects are recognized seems plausible and realistic. Moreover, the reported estimates of FH use are not only low, but also rather stable across a variety of procedural variations that could be considered problematic to a proper assessment of FH use. More precisely, assessing retrieval fluency only at the end of the experiment (and not at the beginning) or presenting the same objects repeatedly during an experiment has little, if any, impact on estimates of FH use.

One feature of the FH niche could be that heuristics such as the FH are more often applied in situations in which effort reduction is warranted, as has been observed for other heuristics. For example, use of the recognition heuristic increases under time pressure (Hilbig, Erdfelder, & Pohl, 2012; Pachur & Hertwig, 2006), deliberative thinking (Hilbig, Scholl, & Pohl, 2010), and depletion of cognitive control (Pohl, Erdfelder, Hilbig, Liebke, & Stahlberg, 2013), and as the cognitive effort of information integration increases (Hilbig, Michalkiewicz, Castela, Pohl, & Erdfelder, 2015).

Note that our results pertain to FH use only—that is, to a simple heuristic that exploits fluency in terms of retrieval latency alone and ignores other knowledge. They say nothing about using fluency in general—for example, as part of other strategies. In alternative decision strategies, fluency might well play a vital role (see, e.g., Hilbig et al., 2011; Marewski et al., 2010; Pohl, 2011). One example is the memory-state heuristic (Erdfelder et al., 2011), which posits that the memory states of the choice objects (as indexed by the speed with which objects are either recognized or rejected as “known”) determine subsequent decision strategies. If one object in a pair is recognized speedily and the other is rejected speedily, reliance on the recognition cue would be most likely. However, if both are recognized or rejected only slowly, guessing or knowledge-based strategies are more likely (see Castela et al., 2014).

In sum, despite the results summarized here, retrieval fluency may well play an important role in heuristic decision making. However, when it serves as the only cue, as assumed

by the FH, fluency appears to have only a minor but nonetheless stable role.

Author note This research was supported by German Research Foundation Grant No. ER 224/2-2 to E.E. and R.F.P. We thank our colleague Herbert Bless, University of Mannheim, for alerting us to the potentially biased estimation of FH use due to specific experimental procedures. We also thank Nikoletta Symeonidou and Sabine Schellhaas for their help in the data analyses.

References

- Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, 13, 219–235. doi:10.1177/1088868309341564
- Anderson, J. R., & Schooler, J. W. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408. doi:10.1111/j.1467-9280.1991.tb00174.x
- Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, 6, 57–86. doi:10.3758/BF03210812
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metammemonic index. *Journal of Experimental Psychology: General*, 127, 55–68. doi:10.1037/0096-3445.127.1.55
- Bröder, A., & Schiffer, S. (2003). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making*, 16, 193–213. doi:10.1002/bdm.442
- Bröder, A., & Schiffer, S. (2006). Adaptive flexibility and maladaptive routines in selecting fast and frugal decision strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 904–918. doi:10.1037/0278-7393.32.4.904
- Brown, N. R., & Tan, S. (2011). Magnitude comparison revisited: An alternative approach to binary choice under uncertainty. *Psychonomic Bulletin & Review*, 18, 392–398. doi:10.3758/s13423-011-0057-1
- Castela, M., & Erdfelder, E. (in press). *The memory state heuristic: A formal model based on repeated judgments*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Castela, M., Kellen, D., Erdfelder, E., & Hilbig, B. E. (2014). The impact of subjective recognition experiences on recognition heuristic use: A multinomial processing tree approach. *Psychonomic Bulletin & Review*, 21, 1131–1138. doi:10.3758/s13423-014-0587-4
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2009). Mix me a list: Context moderates the truth effect and the mere-exposure effect. *Journal of Experimental Social Psychology*, 45, 1117–1122. doi:10.1016/j.jesp.2009.06.019
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth effect: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, 14, 238–257. doi:10.1177/108886830932251
- Erdfelder, E., Auer, T.-S., Hilbig, B. E., Aßfalg, A., Moshagen, M., & Nadarevic, L. (2009). Multinomial processing tree models: A review of the literature. *Zeitschrift für Psychologie/Journal of Psychology*, 217, 108–124. doi:10.1027/0044-3409.217.3.108
- Erdfelder, E., Küpper-Tetzl, C. E., & Mattern, S. D. (2011). Threshold models of recognition and the recognition heuristic. *Judgment and Decision Making*, 6, 7–22.
- Fiedler, K. (2010). How to study cognitive decision algorithms: The case of the priority heuristic. *Judgment and Decision Making*, 5, 1–12.
- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review*, 103, 650–669. doi:10.1037/0033-295X.103.4.650
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109, 75–90. doi:10.1037/0033-295X.109.1.75
- Greifeneder, R., Bless, H., & Pham, M. T. (2010). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review*, 15, 107–141. doi:10.1177/1088868310367640
- Hansen, J., & Wänke, M. (2013). Fluency in context: What makes processing experiences informative. In R. Greifeneder & C. Unkelbach (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behavior* (pp. 70–84). London, UK: Psychology Press.
- Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, 7, 185–207. doi:10.3758/BF03212979
- Hertwig, R., Herzog, S. M., Schooler, L. J., & Reimer, T. (2008). Fluency heuristic: A model of how the mind exploits a by-product of information retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 1191–1206. doi:10.1037/a0013025
- Hertwig, R., Pachur, T., & Kurzenhäuser, S. (2005). Judgments of risk frequencies: Tests of possible cognitive mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 621–642. doi:10.1037/0278-7393.31.4.621
- Herzog, S. M., & Hertwig, R. (2013). The ecological validity of fluency. In C. Unkelbach & R. Greifeneder (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behaviour* (pp. 190–219). Hove, U.K.: Psychology Press.
- Hilbig, B. E. (2010). Reconsidering “evidence” for fast-and-frugal heuristics. *Psychonomic Bulletin & Review*, 17, 923–930. doi:10.3758/PBR.17.6.923
- Hilbig, B. E. (2012). Good things don’t come easy (to mind): Explaining framing effects in judgments of truth. *Experimental Psychology*, 59, 38–46. doi:10.1027/1618-3169.a000124
- Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2011). Fluent, fast, and frugal? A formal model evaluation of the interplay between memory, fluency, and comparative judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 827–839. doi:10.1037/a0022638
- Hilbig, B. E., Erdfelder, E., & Pohl, R. F. (2012). A matter of time: Antecedents of one-reason decision making based on recognition. *Acta Psychologica*, 141, 9–16. doi:10.1016/j.actpsy.2012.05.006
- Hilbig, B. E., Michalkiewicz, M., Castela, M., Pohl, R. F., & Erdfelder, E. (2015). Whatever the cost? Information integration in memory-based inferences depends on cognitive effort. *Memory & Cognition*, 43, 659–671. doi:10.3758/s13421-014-0493-z
- Hilbig, B. E., & Pohl, R. F. (2008). Recognizing users of the recognition heuristic. *Experimental Psychology*, 55, 394–401. doi:10.1027/1618-3169.55.6.394
- Hilbig, B. E., & Pohl, R. F. (2009). Ignorance- versus evidence-based decision making: A decision time analysis of the recognition heuristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1296–1305. doi:10.1037/a0016565
- 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.
- Hu, X., & Batchelder, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, 59, 21–47. doi:10.1007/BF02294263
- Koriat, A. (2008). Easy comes, easy goes? The link between learning and remembering and its exploitation in metacognition. *Memory & Cognition*, 36, 416–428. doi:10.3758/MC.36.2.416
- Lloyd, M. E., Westerman, D. L., & Miller, J. K. (2003). The fluency heuristic in recognition memory: The effect of repetition. *Journal of Memory and Language*, 48, 603–614. doi:10.1016/S0749-596X(02)00535-1

- Marewski, J. N., Gaissmaier, W., Schooler, L. J., Goldstein, D. G., & Gigerenzer, G. (2010). From recognition to decisions: Extending and testing recognition-based models for multialternative inference. *Psychonomic Bulletin & Review*, 17, 287–309. doi:10.3758/PBR.17.3.287
- Marewski, J. N., & Mehlhorn, K. (2011). Using the ACT-R architecture to specify 39 quantitative process models of decision making. *Judgment and Decision Making*, 6, 439–519.
- Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, 118, 393–437. doi:10.1037/a0024143
- Martignon, L., & Hoffrage, U. (1999). Why does one-reason decision making work? A case study in ecological rationality. In G. Gigerenzer, P. M. Todd, & A. R. Group (Eds.), *Simple heuristics that make us smart* (pp. 119–140). New York, NY: Oxford University Press.
- Michalkiewicz, M., Arden, K., & Erdfelder, E. (2016). *Do smarter people make better decisions? The influence of intelligence on adaptive use of the recognition heuristic*. Manuscript submitted for publication.
- Michalkiewicz, M., & Erdfelder, E. (2016). Individual differences in use of the recognition heuristic are stable across time, choice objects, domains, and presentation formats. *Memory & Cognition*, 44, 454–468. doi:10.3758/s13421-015-0567-6
- Michalkiewicz, M., Minich, B., & Erdfelder, E. (2016). *Explaining individual differences in fast-and-frugal decision making: The impact of need for cognition and faith in intuition on use of the recognition heuristic*. Manuscript submitted for publication.
- Moshagen, M. (2010). MultiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, 42, 42–54.
- Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language*, 70, 1–12. doi:10.1016/j.jml.2013.09.007
- Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision making: The success of "success". *Journal of Behavioral Decision Making*, 17, 117–137. doi:10.1002/bdm.465
- Newell, B. R., & Shanks, D. R. (2007). Recognising what you like: Examining the relation between the mere-exposure effect and recognition. *European Journal of Cognitive Psychology*, 19, 103–118. doi:10.1080/09541440500487454
- Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: Retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 983–1002. doi:10.1037/0278-7393.32.5.983
- Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, 19, 251–271. doi:10.1002/bdm.522
- Pohl, R. F. (2011). Recognition information in inferential decision making: An overview of the debate. *Judgment and Decision Making*, 6, 423–438.
- Pohl, R. F., Erdfelder, E., Hilbig, B. E., Liebke, L., & Stahlberg, D. (2013). Effort reduction after self-control depletion: The role of cognitive resources in use of simple heuristics. *Journal of Cognitive Psychology*, 25, 267–276. doi:10.1080/20445911.2012.758101
- Pohl, R. F., & Hilbig, B. E. (2012). The role of subjective linear orders in probabilistic inferences. *Psychonomic Bulletin & Review*, 19, 1178–1186. doi:10.3758/s13423-012-0289-8
- Poldrack, R. A., & Logan, G. D. (1997). Fluency and response speed in recognition judgments. *Memory and Cognition*, 25, 1–10. doi:10.3758/BF03197280
- Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring "fringe" consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, 13, 47–60. doi:10.1016/S1053-8100(03)00049-7
- Sanchez, C. A., & Jaeger, A. J. (2015). If it's hard to read, it changes how long you do it: Reading time as an explanation for perceptual fluency effects on judgment. *Psychonomic Bulletin & Review*, 22, 206–211. doi:10.3758/s13423-014-0658-6
- Scholl, S. G., Greifeneder, R., & Bless, H. (2014). When fluency signals truth: Prior successful reliance on fluency moderates the impact of fluency on truth judgments. *Journal of Behavioral Decision Making*, 27, 268–280. doi:10.1002/bdm.1805
- Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, 112, 610–628. doi:10.1037/0033-295X.112.3.610
- Schwarz, N. (2004). Metacognitive experiences in consumer judgment and decision making. *Journal of Consumer Psychology*, 14, 332–348. doi:10.1207/s15327663jcp1404_2
- Schweickart, O., & Brown, N. R. (2014). Magnitude comparison extended: How lack of knowledge informs comparative judgments under uncertainty. *Journal of Experimental Psychology: General*, 143, 273–294. doi:10.1037/a0031451
- Schwikert, S. R., & Curran, T. (2014). Familiarity and recollection in heuristic decision making. *Journal of Experimental Psychology: General*, 143, 2341–2365. doi:10.1037/xge0000024
- Susser, J. A., Jin, A., & Mulligan, N. W. (2016). Identity priming consistently affects perceptual fluency but only affects metamemory when primes are obvious. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42, 657–662. doi:10.1037/xlm0000189
- Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition*, 43, 647–658. doi:10.3758/s13421-014-0479-x
- Unkelbach, C., & Greifeneder, R. (2013). A general model of fluency effects in judgment and decision making. In C. Unkelbach & R. Greifeneder (Eds.), *The experience of thinking: How the fluency of mental processes influences cognition and behaviour* (pp. 11–32). Hove, UK: Psychology Press.
- Volz, K. G., Schooler, L. J., & von Cramon, D. Y. (2010). It just felt right: The neural correlates of the fluency heuristic. *Consciousness and Cognition*, 19, 829–837. doi:10.1016/j.concog.2010.05.014
- von Helversen, B., Gendolla, G. H. E., Winkelmann, P., & Schmidt, R. E. (2008). Exploring the hardship of ease: Subjective and objective effort in the ease-of-processing paradigm. *Motivation and Emotion*, 32, 1–10. doi:10.1007/s11031-008-9080-6
- Wänke, M., & Hansen, J. (2015). Relative processing fluency. *Psychological Science*, 24, 195–199. doi:10.1177/0963721414561766
- Westerman, D. L., Miller, J. K., & Lloyd, M. E. (2003). Change in perceptual form attenuates the use of the fluency heuristic in recognition. *Memory & Cognition*, 31, 619–629.
- Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language*, 49, 62–79. doi:10.1016/S0749-596X(03)00009-3
- Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 547–565. doi:10.1037/0278-7393.26.3.547