Mem Cogn
DOI 10.3758/s13421-016-0685-9

Task difficulty moderates the revelation effect

André Aßfalg[1,2] · **Devon Currie**[1,3] · **Daniel M. Bernstein**[1]

Abstract Tasks that precede a recognition probe induce a more liberal response criterion than do probes without tasks—the "revelation effect." For example, participants are more likely to claim that a stimulus is familiar directly after solving an anagram, relative to a condition without an anagram. Revelation effect hypotheses disagree whether hard preceding tasks should produce a larger revelation effect than easy preceding tasks. Although some studies have shown that hard tasks increase the revelation effect as compared to easy tasks, these studies suffered from a confound of task difficulty and task presence. Conversely, other studies have shown that the revelation effect is independent of task difficulty. In the present study, we used new task difficulty manipulations to test whether hard tasks produce larger revelation effects than easy tasks. Participants (N = 464) completed hard or easy preceding tasks, including anagrams (Exps. 1 and 2) and the typing of specific arrow key sequences (Exps. 3–6). With sample sizes typical of revelation effect experiments, the effect sizes of task difficulty on the revelation effect varied considerably across experiments. Despite this variability, a consistent data pattern emerged: Hard tasks produced larger revelation effects than easy tasks. Although the present study falsifies certain revelation effect hypotheses, the general vagueness of revelation effect hypotheses remains.

Keywords Recognition · Cognitive illusion · Revelation effect

Judgments can change on the basis of the activities directly preceding them. In the context of recognition experiments, Watkins and Peynircioglu (1990) discovered that revealing a word letter by letter (e.g., *_p_ _ _* → *_p_l_* → *ap_l_* → *ap_le* → *apple*) increases the perceived familiarity of the revealed word relative to a word presented intact (see Aßfalg, 2017, for an overview). This *revelation effect* also occurs when a task preceding the judgment is unrelated to the recognition probe (Westerman & Greene, 1996). For example, directly after solving a simple addition problem (e.g., *256 + 412 = ?*), participants are more likely to claim that a word (e.g., *apple*) appears familiar compared to a word preceded by no addition problem (Niewiadomski & Hockley, 2001). In general terms, the revelation effect represents an increase in "old" judgments following a preceding task (*task condition*), relative to a condition without a preceding task (*no-task condition*).

Criterion shifts and task difficulty in recognition memory

Recognition memory is commonly described in terms of the participant's sensitivity—the ability to discriminate old from new items—and the participant's criterion—the point above which an item appears familiar enough to be called "old" (Macmillan & Creelman, 2005). The revelation effect represents a within-list criterion shift: In a list of randomly intermixed task and no-task trials, the criterion is more liberal

Portions of this research were presented at the 2014 meeting of the Psychonomic Society and the 2015 meeting of the Society for Applied Research in Memory and Cognition.

✉ André Aßfalg
andre.assfalg@psychologie.uni-freiburg.de

[1] Department of Psychology, Kwantlen Polytechnic University, Surrey, BC, Canada

[2] Present address: Department of Psychology, Albert-Ludwigs University, Freiburg, Engelbergerstr. 41, 79085 Freiburg, Germany

[3] Present address: University of Calgary, Calgary, AB, Canada

Published online: 27 December 2016

in the task trials as compared to the no-task trials. Subsequent to the discovery of the revelation effect, other within-list criterion shifts have been identified. For example, in Rhodes and Jacoby's (2007) study, participants provided recognition judgments for words that appeared either on the left or the right side of the screen. Unknown to participants, the words on one side of the screen were mostly studied words and the words on the other side of the screen were mostly unstudied words. In this situation, a criterion shift appeared: The participants' criterion was relatively liberal for the side of the screen with mostly studied words and relatively conservative for the side of the screen with mostly unstudied words. Similarly, Singer and Wixted (2006) found within-list criterion shifts when the retention interval varied across test items: Items tested immediately after study received a more conservative criterion than items studied two days before. Surprisingly, the within-list criterion shifts disappeared when Singer and Wixted compared immediate testing with a 40-min delay—still a strong experimental manipulation. Unlike these within-list criterion shifts, the revelation effect occurs after brief preceding tasks such as solving an anagram. In our experiments, participants typically solve an anagram in about 15 s.

To explain the revelation effect, Hicks and Marsh (1998) argued that participants apply a more liberal criterion to the hard task condition than in the relatively easy no-task condition. This argument partially rests on work by Hirshman (1995) who investigated the impact of list-strength manipulations on criteria. For example, Hirshman found a more liberal criterion for weak items in a pure list of weak items than for weak items in a mixed list of weak and strong items. Because participants should have worse memory for a pure list of weak items than for a mixed list of weak and strong items, this outcome suggests that hard recognition tasks produce more liberal criteria than easy tasks. Similarly, Hirshman found a more liberal criterion for strong items in a mixed list of weak and strong items than for strong items in a pure list of strong items. Because participants should have worse memory for a mixed list of weak and strong items than for a pure list of strong items, this outcome again suggests that hard recognition tasks produce more liberal criteria than easy tasks. The hypothesis that hard tasks produce liberal criteria also predicts the more liberal criterion for test items after a two-day delay than immediately after testing in Singer and Wixted's (2006) study. After a two-day delay items are harder to recognize than after immediate testing, again producing a more liberal criterion.

Task difficulty and the revelation effect

The idea that the difficulty of the preceding task induces a liberal criterion as compared to the no-task condition is part of the *decrement-to-familiarity hypothesis* (Hicks & Marsh, 1998). Hicks and Marsh proposed that the preceding task activates memory traces that are related to the stimulus in the preceding task. This activation is assumed to persist at least until the recognition judgment, in which the activated memory traces compete with the recognition probe. This competition is further assumed to increase the difficulty of the recognition judgment, and participants react to this increased difficulty by applying a more liberal criterion than in the no-task condition. From these assumptions it is straightforward to derive the prediction that hard preceding tasks should be more likely to produce a liberal criterion in the task condition compared to easy preceding tasks. Thus, hard preceding tasks should also be more likely than easy preceding tasks to produce a revelation effect.

Hard tasks should also produce a larger revelation effect than easy tasks according to the *discrepancy-attribution hypothesis* (Whittlesea & Williams, 1998, 2000, 2001), which states that participants perceive the preceding task as relatively harder than the judgment. Hard tasks require mental effort, which participants constantly track metacognitively as varying levels of fluency (Oppenheimer, 2008). Fluency—the ease and speed of information processing—serves as a heuristic cue for judgments such as recognition, truth, or aesthetical pleasure (Alter & Oppenheimer, 2009; Jacoby & Dallas, 1981; Reber, Schwarz, & Winkielman, 2004). These judgments relate to fluency in everyday life. For example, a stimulus that a person often encounters will increase the fluency with which this person processes the stimulus in the future. Recognition judgments that rely on fluency will therefore tend to be accurate. However, because of the close link between fluency and several judgment types, fluency can be misattributed to the wrong source. For example, participants judge masked words as less familiar than unmasked words, arguably because the masking reduces the fluency with which participants process those words (Whittlesea, Jacoby, & Girard, 1990).

The discrepancy-attribution hypothesis further specifies that not all fluency will cause misattribution. Instead, observed fluency must be discrepant from expected fluency to cause misattribution. For example, Whittlesea and Williams (1998) found that participants judge easy-to-pronounce nonwords (e.g., *hension*) as more familiar than hard-to-pronounce nonwords (e.g., *jufict*). Arguably, for nonwords, expected fluency is lower than for words. However, pronounceable nonwords violate the expectation of low fluency, thus causing a misattribution of fluency to familiarity with the nonword. Some have argued that, in revelation effect experiments, fluency discrepancy occurs between the preceding task and the recognition probe (Bernstein, Whittlesea, & Loftus, 2002; Whittlesea & Williams, 2001): When participants encounter the relatively hard preceding task, they set an expectation of low fluency that is discrepant with the relatively high fluency with which they process the recognition probe. On the basis of

this assumption, relative to easy preceding tasks, hard preceding tasks should be more likely to cause a discrepancy and, therefore, a revelation effect.[1]

However, not all hypotheses suggest that task difficulty moderates the revelation effect. According to the *criterion-flux hypothesis*, the preceding task disrupts working memory, temporarily removing important test-list information such as the proportion of studied items (Hockley & Niewiadomski, 2001; Niewiadomski & Hockley, 2001). Hockley and Niewiadomski suggested that participants apply a liberal default criterion under these circumstances. Until the next trial begins, participants restore the original conservative criterion. Thus, the criterion is in constant flux between a liberal criterion in the task condition and a conservative criterion in the no-task condition. Niewiadomski and Hockley (2001) found equally sized revelation effects when the preceding task included a single anagram as compared to two anagrams in quick succession. The authors concluded that "Two problem tasks would not generally result in a larger effect than only one preceding task, because, usually, one task would be sufficient to displace list context from working memory" (Niewiadomski & Hockley, 2001, p. 1137). Thus, according to the criterion-flux hypothesis, there is a low threshold for the disruption of working memory that, once crossed, will trigger a revelation effect. This suggests that manipulations of task difficulty are unlikely to moderate the size of the revelation effect.

Similarly, the *global-matching hypothesis* suggests that task difficulty does not moderate the size of the revelation effect if the preceding task and study-list items are dissimilar (Westerman & Greene, 1998). According to this hypothesis, the preceding task increases the activation of study-list memory traces. Following the rationale of global-matching models of memory, this additional activation is assumed to contribute to the experience of familiarity (e.g., Clark & Gronlund, 1996). Consequently, participants provide more "old" responses directly after the preceding task than without. In principle, a difficult preceding task could activate more memory traces than an easy task. However, the hypothesis also predicts that the preceding task has to resemble the study-list items in order to produce a revelation effect. When Westerman and Greene did not find a revelation effect for verbal material in the study list and a math problem as preceding task, they concluded that the math problem did not sufficiently activate study-list memory traces. In most of the present experiments, the study-list items and the preceding task are unrelated. Thus, according to the global-matching hypothesis, we should not observe a revelation effect in these experiments. Consequently, task difficulty should not moderate the revelation effect.

[1] Note that the discrepancy - attribution hypothesis does not specify whether the revelation effect is continuous or discrete. However, even if the revelation effect appears in an all-or-nothing fashion, it is possible that the revelation effect occurs for fewer participants in the easy condition than for the hard condition. Thus, on the level of aggregated recognition judgments, the revelation effect could be present in the easy condition but smaller than for the hard condition.

Empirical evidence for the effect of task difficulty on the revelation effect

Although some of the aforementioned revelation effect hypotheses predict a positive correlation between task difficulty and the size of the revelation effect, the empirical evidence mostly contradicts this prediction. Peynircioğlu and Tekcan (1993) found no correlation between anagram solution times and the tendency to respond "old" in a recognition judgment directly following the anagram. Further, several studies have manipulated task difficulty by varying the stimuli in the preceding task. For example, the revelation effect is equally large when participants solve five- versus eight-letter anagrams, transpose the order of elements in three- versus eight-digit numbers, or perform three- versus eight-letter memory-span tasks (Verde & Rotello, 2003; Watkins & Peynircioglu, 1990; Westerman & Greene, 1998). Arguably, preceding tasks with longer relative to shorter stimuli are harder to solve and, according to several revelation effect hypotheses, should have elicited a larger revelation effect.

Other studies have found higher frequency judgments and more recognition claims with increasing stimulus length in the task condition (Bornstein & Neely, 2001; Watkins & Peynircioglu, 1990) and this has been taken as evidence that hard tasks produce a larger revelation effect than easy tasks (Niewiadomski & Hockley, 2001). However, these studies confounded the stimulus length manipulation and the task presence manipulation (task vs. no task): The effect of stimulus length was analyzed with the no-task condition serving as a zero-length preceding task. Although the descriptive results in Bornstein and Neely's (2001) experiments suggest a positive correlation between the size of the revelation effect and task difficulty, the authors do not report whether this effect is statistically reliable.

Niewiadomski and Hockley (2001) took a different approach to determine the influence of task difficulty on the size of the revelation effect. These authors hypothesized that task difficulty increases not with the length of the stimulus in the preceding task but with the number of tasks directly preceding the judgment. Thus, in several experiments, participants either solved a single anagram, a single addition problem, or two consecutive tasks consisting of a combination of anagrams and addition problems. However, the size of the revelation effect was independent of the number and type of tasks performed. Thus, these experiments showed that task difficulty did not moderate the revelation effect.

The current data are inconclusive regarding task difficulty's role in the revelation effect. We believe that this is partly due to the choice of task difficulty manipulations in extant research. Take manipulations of task difficulty, for example, that

change the length of the task stimulus (e.g., five- vs. eight-letter anagrams). It is not clear that it is harder to search for the solution to an eight-letter anagram than to a five-letter anagram. It is even possible that eight-letter anagrams are easier to solve than five-letter anagrams, if participants have more experience with eight- than with five-letter words. An arguably better approach would be to experimentally control whether participants practice a task.

A final concern regarding the aforementioned studies is statistical power. Most of these experiments manipulated task difficulty within subjects with samples sizes of 21 to 44 participants (Bornstein & Neely, 2001; Verde & Rotello, 2003; Watkins & Peynircioglu, 1990; Westerman & Greene, 1998). Similarly, in a between-subjects design, Niewiadomski and Hockley (2001) had between 40 and 45 participants per group. The mostly negative results in these studies suggest that any moderating effect of task difficulty on the revelation effect is likely small and requires larger sample sizes.

The present study

We presented participants with relatively hard versions of common (Exps. 1 and 2) and novel preceding tasks (Exps. 3–6). We contrasted these hard tasks with highly similar, but easier versions of the same tasks that allowed participants to rely on well-practiced skills. In all our experiments, participants received task and no-task trials within-subjects. However, we manipulated the task difficulty between subjects: In the hard and easy conditions, respectively, participants received a hard or easy version of the preceding task. In Experiments 1 and 2, participants either solved a new anagram in each task trial (hard condition) or received the same anagram repeatedly (easy condition). Solving an anagram is the most common preceding task in revelation-effect research (e.g., Aßfalg & Nadarevic, 2015). In Experiments 3 and 4, we introduced a novel preceding task in which participants pressed a random sequence of arrow keys on a computer keyboard (hard condition) or pressed the same arrow key repeatedly (easy condition). We repeated this procedure in Experiments 5 and 6, but here participants always pressed sequences of two keys that were either random (hard condition) or included a fixed sequence (easy condition).

Note that hard and easy versions of the preceding tasks in Experiments 1–6 were identical in almost all respects. Hard and easy versions of the preceding task required similar cognitive processes and motor responses. The critical difference is that, as compared to hard versions, easy versions repeated stimuli and responses, allowing practice not only with the general characteristics of the task but also with particular stimulus–response pairs.

Further, assuming that any moderating effect of task difficulty on the revelation effect would be small, we decided to increase the overall sample size as compared to similar studies. Instead of a single large-sample experiment, we decided to collect data for multiple experiments with relatively smaller sample sizes and to combine the result of these experiments afterward to increase overall power. This approach is partially the result of our decision to use a mixture of preceding tasks to ensure that any effects would not rely on a single preceding task. Thus, Experiments 1, 3, and 5 included three different preceding tasks. We were also interested in the replicability of our results. To that end, Experiments 2, 4, and 6 were direct replications of Experiments 1, 3, and 5, respectively. Recently, the replicability of results in psychological studies has attracted much attention (Open Science Collaboration, 2015). Part of the issue concerns the large degree of uncertainty associated with results in underpowered studies. On the basis of the mixed findings in the literature, we suspected that this might be an issue in revelation effect research as well. We hoped to inform our own research as well as that of others by illustrating the changes in effect sizes, even in direct replications, with sample sizes typical of the literature.

Predictions for Experiments 1–6

On the basis of the familiarity-decrement and discrepancy-attribution hypotheses, we expected that hard preceding tasks would produce a larger revelation effect than easy tasks. For example, according to the discrepancy-attribution hypothesis, participants should process difficult preceding tasks less fluently than the recognition probe. This fluency discrepancy should further be misattributed to familiarity with the probe, causing a revelation effect. Conversely, easy preceding tasks should be processed (nearly) as fluently as the recognition probe, thus reducing the chance of discrepancy and the revelation effect. Conversely, the criterion-flux and global-matching hypotheses predict no effect of task difficulty on the size of the revelation effect. The global-matching hypothesis further predicts the absence of any revelation effect in experiments with the arrow-key task (Exps. 3–6). Furthermore, to establish that the revelation effect depends on task difficulty, the revelation effect should be larger in the hard than in the easy condition and the effect in the easy condition should be statistically significant. This requirement ensures that the easy condition is not just functionally equivalent to having no task at all but includes a genuine but easy preceding task.

Method

Power analysis

We performed power analysis with the software G*Power (Faul, Erdfelder, Buchner, & Lang, 2009). In each experiment, we were interested in the effect of task presence (task vs. no

task)—a within-subjects manipulation—between two participant groups (e.g., hard task vs. easy task). Thus, our analyses focused on the detection of a within–between interaction. To detect a medium-sized effect (f = 0.25; see Cohen, 1992), assuming $\alpha = \beta = .05$ and a medium-sized correlation between judgments in the task and no-task conditions of r = .50, the required sample size was 27 per participant group. We chose a conservative approach and exceeded this sample size in Experiments 1–6 (see Table 1). Note that Experiments 1–6 included sample sizes per group that were equal to or in excess of the sample sizes in other studies that have investigated the influence of task difficulty on the revelation effect (Niewiadomski & Hockley, 2001; Verde & Rotello, 2003; Watkins & Peynircioglu, 1990; Westerman & Greene, 1998).

Participants

Table 1 lists the demographic data of our participant samples in Experiments 1–6. All participants were US residents recruited via online crowdsourcing sites. Each participant received $0.50 (USD) for the 20 min it took to complete the experiment. We did not exclude any participants from the analyses.

Material

We used a pool of 130 English eight-letter words retrieved from the MRC psycholinguistic database (Wilson, 1988) as the stimuli in Experiments 1–6. The words had a written frequency of more than 50 per million (Kučera & Francis, 1967).

Design

All of the experiments included a 2 (Item Status: old vs. new) × 2 (Task Difficulty: hard vs. easy) × 2 (Task Presence: task vs. no task) design with Task Difficulty as the only between-subjects factor. Our only dependent variable was the recognition confidence judgment, which we also analyzed in terms of criterion estimates.

Table 1 Demographic data in Experiments 1–6

Experiment	*N*			Gender			Age	
	Hard	Easy	Total	Female	Male	Missing	*M*	*SD*
1	42	42	84	42	37	5	29.42	10.03
2	25	35	60	40	18	2	34.43	11.66
3	40	42	82	49	32	1	34.22	11.72
4	40	41	81	48	26	7	36.61	12.96
5	42	42	84	59	22	3	33.35	12.43
6	34	39	73	50	23	0	35.40	14.69
Total	223	241	464	288	158	18	33.85	12.46

General procedure

Participants provided informed consent and completed the experiment online at a computer of their choice. Each experiment began with a study phase, in which 40 words appeared sequentially on the center of the screen for 2 s each. Of these 40 study words, all participants received the same five words as primacy buffers and the same five words as recency buffers.

After the study phase, all participants practiced the preceding task. The preceding task varied across Experiments 1–6 and is described in the next section. All participants entered the test phase of the experiment after practicing the preceding task. The test phase of each experiment included 30 old words from the study phase randomly intermixed with 30 new words. Critically, to induce a revelation effect, the preceding task appeared before half the recognition probes. After the preceding task, the recognition probe immediately appeared above a 6-point response scale ranging from *sure new* to *sure old*. Once participants provided their recognition confidence judgment, a new trial started automatically. Conversely, in the no-task condition, words appeared without a preceding task and participants simply provided a recognition confidence judgment using the same response scale as in the task condition. The computer randomized which words appeared as old words, new words, and whether a word appeared in the task condition or the no-task condition for each participant anew. After the participants had finished the test phase, they received debriefing and payment.

Preceding tasks in Experiments 1–6

In the preceding task of Experiments 1 and 2, participants solved anagrams. Each anagram appeared on the center of the screen. Below each anagram, the computer presented a solution code. For example, the anagram *esnnuhis* was accompanied by the solution code 84732561. The computer instructed participants that the letter directly above the "1" in the solution code would be the first letter of the anagram solution, the letter above the "2" would be the second letter, and so on. In the present example, this algorithm leads to the solution *sunshine*. In the hard condition, participants practiced the preceding task with the five primacy buffers. Participants later solved a different, previously unpresented, anagram on each trial of the test phase. Thus, the hard condition followed the procedure in a typical revelation effect experiment (e.g., Aßfalg & Nadarevic, 2015). Conversely, in the easy condition participants practiced the preceding task with a randomly chosen word that was not part of the study and test lists. The same word appeared later on every trial in the preceding task of the test phase. There was no time limit for solving anagrams, and each anagram stayed on the screen until participants typed the correct solution.

In the preceding task of Experiments 3 and 4, participants typed arrow key sequences. In this task, a red rectangle appeared in the upper half of the screen, centered on the horizontal axis. The computer displayed a sequence of symbols representing the arrow keys on a keyboard. Each symbol started in the lower half of the screen and gradually moved through the red rectangle toward the upper edge of the screen. The symbols appeared at a rate of approximately one symbol per second. The participants' task was to press the arrow key corresponding to the symbol currently appearing inside the red rectangle. In the hard condition, the computer randomly determined for each symbol whether it would show the up, down, left, or right arrow key (Fig. 1A). Conversely, in the easy condition, all symbols showed the up arrow key (Fig. 1B). The computer provided accuracy feedback to the participants by displaying the initial grayscale symbol in yellow, after a correct keypress, or in red, after an incorrect or missing keypress. Symbols reaching the upper end of the screen disappeared automatically. All participants practiced the arrow-key task prior to entering the test phase of the experiment. For each participant in the preceding task, practice ended automatically after 2 min, or once the computer registered 20 correct keypresses. In each task trial of the test phase, the preceding task appeared for 10 s. This roughly equaled the average anagram solution time in the hard condition of Experiments 1 and 2. Directly afterward, the recognition probe appeared automatically and the participants provided their recognition judgments.

The preceding task in Experiments 5 and 6 was identical to that in Experiments 3 and 4, with the following exceptions. We did not vary the number of arrow keys. Instead, the participants in the hard condition received a randomized arrow key sequence involving only the left and right arrow keys. This sequence was newly randomized in each task trial, making repetitions of the same sequence highly unlikely. Conversely, participants in the easy condition received a repeating sequence involving the left and right arrow keys: left, left, right, right, left, left, right, right, and so on. The same sequence appeared for 15 s in all task trials and in the practice phase preceding the recognition test. We increased the duration of the preceding task as compared to those of Experiments 3 and 4 to emphasize the differences between the random and fixed arrow key sequences. We chose the left–left–right–right repeating sequence in the easy condition to ensure that the probability of a switch from the left to the right arrow key, and vice versa, was on average .5 for both the hard and easy conditions.

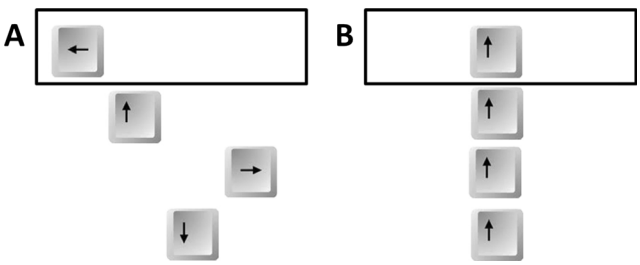

Fig. 1 Example illustrating the preceding task in Experiments 3 and 4. A sequence of symbols representing a random sequence of all four arrow keys (A) or the up arrow key only (B) moved from the lower half of the screen upward through a rectangle. The participants' task was to press the arrow key currently appearing in the rectangle. Panel A depicts the hard preceding task, in which participants had to press a random sequence of the four arrow keys. Conversely, panel B depicts the easy preceding task, in which participants pressed the up arrow key in all task trials. Note that in the hard condition (A), the left, up, down, and right arrow keys appeared in fixed horizontal positions—from left to right: left key, up key, down key, and right key

Results[2]

Preceding task performance

Table 2 lists the participants' performance in the preceding tasks of Experiments 1–6. For Experiments 1 and 2, we computed anagram solution times in the hard and easy conditions as a manipulation check of task difficulty. To mitigate the effect of solution time outliers, we computed the median instead of the mean solution times per participant. In both experiments, the anagram solution time was considerably slower in the hard than in the easy condition (see Table 2). To assess the participants' performance in the arrow key task of Experiments 3–6, we computed the average proportions of correct responses for each participant across all task trials. In most experiments, participants in the hard condition managed fewer correct responses than did participants in the easy condition. However, this difference was only significant in Experiment 4. Overall, the participants had little trouble pressing the correct arrow keys. This ceiling effect likely limits the usefulness of the proportions of correct responses as a proxy for task difficulty.

The revelation effect

We assessed the revelation effect on the level of recognition-confidence judgments and within the framework of signal detection theory (SDT). To determine the preceding task's effect on the SDT measures in Experiments 1–6, we computed each participant's criterion $c_a = (-2^{0.5}s/[(1 + s^2)^{0.5}(1 + s)])[z(H) + z(F)]$, where s is the slope of the participant's receiver operating characteristic in z-space, z is the inverse of the standard-normal distribution function, and H and F are the participant's hit and false-alarm rates (Macmillan & Creelman, 2005).[3] Although it was not critical for present

[2] The raw data of Experiments 1–6 are available at osf.io/uwrgp.

[3] The slope could not be estimated for participants who gave the same response in a condition, such as exclusively responding "sure old" to all new items in the task condition. Consequently, SDT measures are not available for all participants.

Table 2 Anagram solution times (Exps. 1 and 2) and proportions of correct arrow keypresses (Exps. 3–6) as a function of task difficulty (hard vs. easy)

Mean (*SD*) Anagram Solution Time (s)			
Experiment	Hard (Different)	Easy (Same)	Effect Size [95% CI]
1	14.10 (5.13)	3.94 (3.17)	d = 2.38, [1.82, 2.94]
2	12.00 (5.19)	2.92 (1.12)	d = 2.63, [1.93, 3.33]
Proportions of Correct Keypresses			
Experiment	Hard (Four)	Easy (One)	Effect Size [95% CI]
3	.90 (.20)	.95 (.17)	d = 0.28, [–0.16, 0.71]
4	.89 (.19)	.98 (.05)	d = 0.61, [0.17, 1.06]
Experiment	Hard (Random)	Easy (Fixed)	Effect Size [95% CI]
5	.93 (.21)	.96 (.15)	d = 0.19, [–0.24, 0.62]
6	.95 (.17)	.91 (.22)	d = –0.21, [–0.67, 0.25]

In Experiments 1 and 2, participants solved different anagrams versus the same anagram across test trials; In Experiments 3 and 4, participants typed sequences consisting of four arrow keys versus one; In Experiments 5 and 6, participants typed random versus fixed sequences of two arrow keys; The reported effect sizes are Cohen's d; Anagram solution times refer to group means based on the participant medians; To complete Experiments 1 and 2, participants had to solve all anagrams

purposes, we also report each participant's sensitivity $d_a = [2/(1 + s)^2]^{0.5}$ $[z(H) - s\ z(F)]$. We computed hit and false-alarm rates by dichotomizing the recognition confidence judgments. Judgments ranging from "guess old" to "sure old" were coded as hits when the noun was old, and as a false alarms when the noun was new. We further applied Snodgrass and Corwin's (1988) correction formula to avoid hit and false-alarm rates of 1 and 0, which would prevent the computation of d_a and c_a. Thus, we added .5 to the hit and false-alarm frequencies and divided by $M + 1$, where M is the number of old or new items.

Table 3 lists the mean recognition-confidence judgments and SDT measures in Experiments 1–6 (see the Appendix for corresponding hit and false-alarm rates). For the sake of completeness, Table 3 also includes sensitivity estimates. However, because the revelation effect is typically observed in recognition-confidence judgments and criterion estimates (Verde & Rotello, 2004), we will not discuss sensitivity any further. The revelation effect occurred in all six experiments: Recognition-confidence judgments were higher in the task condition than in the no-task condition in all experiments (all Fs ≥ 6.83, all ps ≤ .011) except Experiment 5, $F(1, 82) = 3.46$, $p = .066$. The absence of a main effect of task presence (task vs. no task) in Experiment 5 can be explained by the significant interaction between task presence and task difficulty (hard vs. easy), $F(1, 82) = 13.99$, $p < .001$, due to a larger revelation effect in the hard condition compared to the easy condition. In all other experiments, task difficulty did not interact with the revelation effect (all Fs ≤ 3.77, all ps ≥ .056). The results for criterion estimates closely mirrored those for recognition confidence judgments. Criterion estimates were more liberal in the task condition than in the no-task condition in all experiments (all Fs ≥ 8.23, all ps ≤ .005)—the revelation effect—with the exception of Experiment 5, $F(1, 82) = 3.79$, $p = .055$. Again, in Experiment 5, the revelation effect was larger in the hard than in the easy condition, $F(1, 82) = 13.19$, $p < .001$. In all other experiments, the interaction between task presence (task vs. no task) and task difficulty (hard vs. easy) was not significant (all Fs ≤ 3.51, all ps ≥ .065).

On the level of individual experiments, and with statistical significance as the only criterion, Experiments 1–6 offer little evidence for a moderating effect of task difficulty on the revelation effect. However, the effect sizes in Table 3 suggest a different conclusion: The revelation effect sizes were consistently larger in the hard than in the easy condition, with the single exception of the criterion estimates in Experiment 4. Figure 2 depicts this data pattern in the effect sizes (Cohen's d) of the revelation effect across Experiments 1–6, separately for task difficulty (hard vs. easy) and response measure (recognition confidence vs. criterion), along with the combined effect sizes across Experiments 1–6 (Borenstein, Hedges, Higgins, & Rothstein, 2009; Cumming, 2012, 2014).[4]

To make full use of the entire data set and to achieve higher statistical power, we combined the data of Experiments 1–6 in an analysis of variance including item status (old vs. new), task difficulty (hard vs. easy), task presence (task vs. no task), and experiment (1–6) as predictors of recognition-confidence judgments and criterion estimates. We included experiment as a predictor in the analysis to test whether the revelation effect or its interaction with task difficulty were mainly driven by a subset of Experiments 1–6. Including the entire data of Experiments 1–6, a revelation effect emerged for recognition confidence judgments, $F(1, 452) = 54.93$, $p < .001$. This revelation effect was further qualified by significant interactions with task difficulty, $F(1, 452) = 16.01$, $p < .001$, and item status, $F(1, 452) = 7.00$, $p = .008$. These interactions indicated a larger effect of task presence for hard than for easy tasks and a larger effect of task presence for new than for old items. Furthermore, these effects occurred independent of the experiment (Fs < 1). The revelation effect for recognition confidence judgments also emerged when we only included the participants from the hard condition, $F(1, 217) = 60.29$, $p < .001$. The revelation effect in the hard condition was independent of the experiment, $F(5, 217) < 1$. Similarly, the revelation effect emerged when we included only the participants from the easy condition, $F(1, 235) = 6.29$, $p = .01$. This revelation effect was qualified by a significant interaction with item status, $F(1, 235) = 5.51$, $p = .02$, indicating a larger effect of task presence for new than for old items. However, the revelation effect in the easy condition was independent of the experiment, $F(5, 235) = 1.48$, $p = .20$.

[4] We computed combined effect sizes with a fixed-effects model, as outlined in Borenstein, Hedges, Higgins, and Rothstein (2009).

Table 3 Mean (*SD*) recognition - confidence judgments and signal detection theory (SDT) measures in Experiments 1–6 as a function of item status (old vs. new), task difficulty (hard vs. easy), and task presence (task vs. no task)

Recognition - Confidence Judgments (1 = *sure new*; 6 = *sure old*)

Experiment/	Difficulty	Old		New		Combined Effect Size [95% CI]
		Task	No Task	Task	No Task	
1	Hard	4.32 (0.69)	4.12 (0.85)	3.54 (0.78)	3.21 (0.91)	0.32 [0.15, 0.50]*
	Easy	4.08 (0.67)	4.10 (0.73)	3.08 (0.97)	2.90 (0.87)	0.08 [–0.09, 0.25]
2	Hard	4.35 (0.76)	4.11 (0.88)	3.18 (0.81)	2.93 (0.90)	0.29 [0.10, 0.48]*
	Easy	4.25 (0.87)	4.16 (0.77)	2.98 (0.93)	2.73 (0.92)	0.19 [–0.01, 0.39]
3	Hard	4.25 (0.74)	4.14 (0.69)	3.18 (0.92)	2.84 (0.85)	0.27 [0.04, 0.49]*
	Easy	4.42 (0.69)	4.34 (0.74)	2.88 (0.78)	2.85 (0.73)	0.08 [–0.09, 0.24]
4	Hard	4.54 (0.87)	4.46 (0.82)	3.01 (0.91)	2.70 (0.83)	0.23 [0.06, 0.39]*
	Easy	4.41 (0.72)	4.31 (0.79)	2.95 (0.86)	2.79 (0.94)	0.15 [–0.02, 0.32]
5	Hard	4.51 (0.71)	4.21 (0.69)	2.89 (0.73)	2.73 (0.74)	0.33 [0.18, 0.48]*
	Easy	4.26 (0.99)	4.41 (0.84)	2.53 (0.81)	2.54 (0.89)	–0.08 [–0.21, 0.04]
6	Hard	4.47 (0.79)	4.20 (0.98)	3.33 (0.82)	3.08 (0.81)	0.29 [0.16, 0.42]*
	Easy	3.94 (0.94)	3.98 (0.79)	3.15 (0.79)	2.98 (0.93)	0.08 [–0.09, 0.25]

SDT Measures

Experiment/	Difficulty	Sensitivity		Criterion		Criterion Effect Size [95% CI]
		Task	No Task	Task	No Task	
1	Hard	0.58 (0.63)	0.54 (0.77)	–0.21 (0.42)	0.01 (0.45)	0.48 [0.23, 0.73]*
	Easy	0.57 (0.66)	0.72 (0.68)	–0.04 (0.47)	0.03 (0.42)	0.15 [–0.10, 0.40]
2	Hard	0.85 (0.83)	0.78 (0.84)	–0.09 (0.35)	0.13 (0.42)	0.55 [0.15, 0.95]*
	Easy	0.84 (0.62)	0.81 (0.65)	–0.02 (0.43)	0.10 (0.44)	0.28 [0.04, 0.51]*
3	Hard	0.72 (0.60)	0.79 (0.64)	–0.11 (0.49)	0.08 (0.41)	0.43 [0.12, 0.73]*
	Easy	0.92 (0.69)	0.98 (0.69)	–0.08 (0.35)	–0.03 (0.46)	0.11 [–0.12, 0.33]
4	Hard	1.06 (0.71)	1.08 (0.58)	–0.16 (0.50)	–0.04 (0.44)	0.26 [0.05, 0.48]*
	Easy	1.07 (0.72)	0.93 (0.72)	–0.09 (0.42)	0.03 (0.48)	0.26 [0.04, 0.49]*
5	Hard	1.06 (0.78)	0.91 (0.68)	–0.11 (0.31)	0.07 (0.31)	0.57 [0.27, 0.88]*
	Easy	1.09 (0.83)	1.10 (0.81)	0.10 (0.37)	0.04 (0.35)	–0.15 [–0.40, 0.10]
6	Hard	0.71 (0.83)	0.82 (0.86)	–0.26 (0.35)	–0.08 (0.44)	0.43 [0.18, 0.68]*
	Easy	0.51 (0.77)	0.59 (0.79)	0.00 (0.45)	0.07 (0.43)	0.15 [–0.14, 0.45]

Effect sizes (Cohen's *d*) refer to the size of the revelation effect (i.e., comparison of task vs. no task) for recognition - confidence judgments and SDT criterion estimates. * $p < .05$

These results were closely mirrored by the criterion estimates. Again, a revelation effect appeared, $F(1, 451) = 58.01$, $p < .001$, which was larger for hard than for easy preceding tasks, $F(1, 451) = 14.55$, $p < .001$. Again, these effects were independent of the experiment (Fs < 1). The revelation effect for criterion estimates also emerged when we only included the participants from the hard condition, $F(1, 216) = 61.28$, $p < .001$. The revelation effect in the hard condition was independent of the experiment, $F(5, 216) < 1$. Similarly, the revelation effect for criterion estimates emerged when we included only the participants from the easy condition, $F(1, 235) = 7.72$, $p = .006$. The revelation effect in the easy condition was again independent of the experiment, $F(5, 235) = 1.44$, $p = .21$.

In terms of the combined effect sizes in Fig. 2, the revelation effect for confidence judgments in Experiments 1–6 was about five times larger in the hard condition ($d = 0.29$, 95% CI = [0.22, 0.36]) than in the easy condition ($d = 0.06$, 95% CI = [–0.01, 0.13]). Similarly, the combined revelation effect for criterion estimates was about three times larger in the hard condition ($d = 0.42$, 95% CI = [0.31, 0.54]) than in the easy condition ($d = 0.14$, 95% CI = [0.04, 0.24]). Thus, the combined analysis of Experiments 1–6 supports the predictions of the familiarity decrement and discrepancy attribution hypotheses that hard tasks elicit a larger revelation effect than do easy tasks.

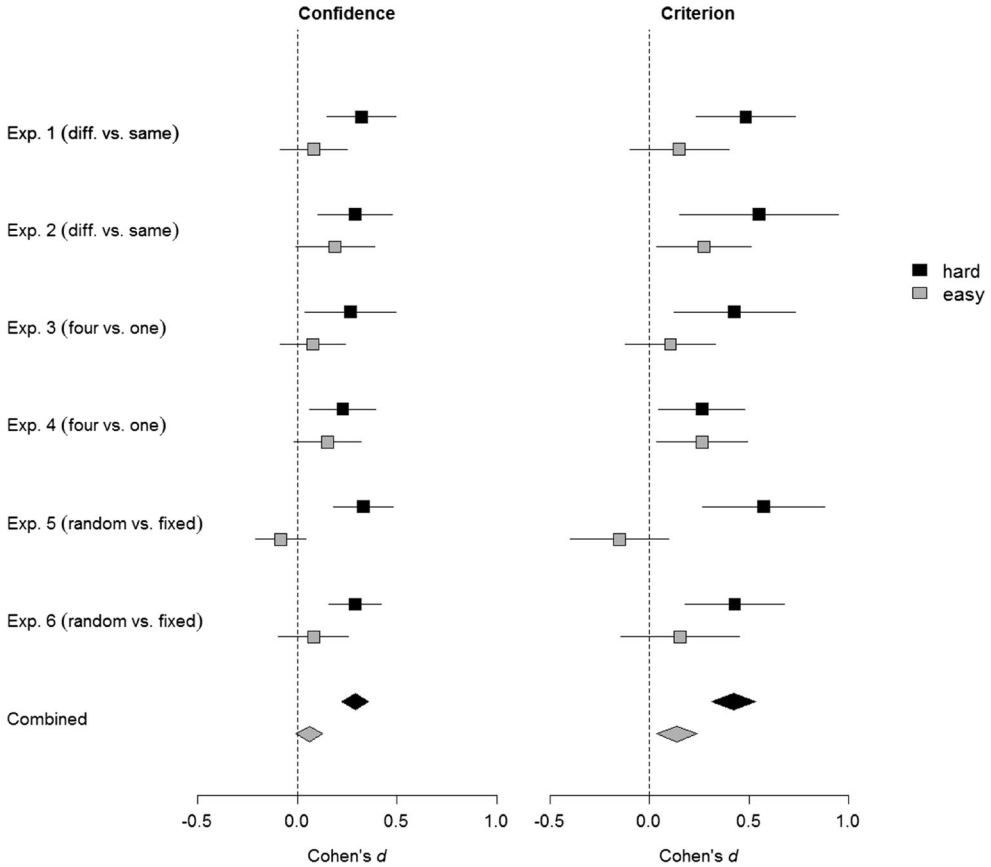

Fig. 2 Effect sizes (Cohen's *d*) of the task presence (task vs. no task) manipulation, shown separately for Experiments 1–6 (squares) and combined across experiments (diamonds), as a function of task difficulty (hard vs. easy) and response measure (confidence vs. criterion). Error bars represent 95% confidence intervals of the individual effect sizes. The width of each diamond symbol represents the 95% confidence interval of the combined effect size. The dotted line marks the null effect. Relative to easy preceding tasks, hard tasks elicited a larger revelation effect. In Experiments 1 and 2, participants solved different anagrams versus the same anagram across test trials (*diff. vs. same*). In Experiments 3 and 4, participants typed sequences of four arrow keys versus one (*four vs. one*). In Experiments 5 and 6, participants typed random versus fixed sequences of two arrow keys (*random vs. fixed*)

Discussion

Some revelation effect hypotheses predict that hard and easy preceding tasks elicit equally large revelation effects. Together, Experiments 1–6 falsified this prediction. In Experiments 1 and 2, participants either solved a different anagram (hard condition) or the same anagram (easy condition) in the preceding task of each task trial. In Experiments 3–6, participants typed random arrow-key sequences (hard condition) or the same arrow-key sequence (easy condition) in each task trial. In all of these experiments, a revelation effect occurred in the hard condition: The participants provided higher recognition confidence judgments in the task condition than in the no-task condition. Critically, in Experiments 1–6, the size of the revelation effect was smaller after an easy than after a hard preceding task.

Previous studies found a moderating effect of task difficulty on the size of the revelation effect (Bornstein & Neely, 2001; Watkins & Peynircioglu, 1990). However, these studies confounded the revelation effect with the task-difficulty manipulation. Without this confound, previous studies failed to find an effect of task difficulty on the revelation effect (Niewiadomski & Hockley, 2001; Verde & Rotello, 2003; Watkins & Peynircioglu, 1990; Westerman & Greene, 1998). On the level of individual experiments, the present experiments also found little evidence of a moderating effect of task difficulty on the revelation effect. An exception was Experiment 5, in which the hard task elicited a larger revelation effect than the easy task. However, when we considered the combined evidence in Experiments 1–6, hard preceding tasks elicited a three to five times larger revelation effect than easy preceding

tasks. Importantly, easy preceding tasks were still sufficient to produce a revelation effect. This suggests that the revelation effect occurs in a gradual rather than an all-or-nothing fashion.

A limitation of the present research concerns the operationalization of task difficulty in Experiments 1–6. Manipulation checks included the time to solve an anagram (Exps. 1 and 2) and the proportion of correct keypresses in the arrow-key task (Exps. 3–6). Whereas the time to solve an anagram indicated that "hard" tasks were indeed more difficult than "easy" tasks, the proportions of correct keypresses suffered from ceiling effects. Similar studies did not report manipulation checks (Niewiadomski & Hockley, 2001; Verde & Rotello, 2003; Watkins & Peynircioglu, 1990; Westerman & Greene, 1998) making a comparison to the present experiments difficult. Despite these difficulties, the manipulation checks in Experiments 1, 2, and 4 successfully demonstrated that our hard tasks were indeed harder than our easy tasks.

Another limitation concerns the experimental paradigm in the present experiments. Researchers have distinguished two experimental paradigms in revelation effect research. In the unrelated paradigm, the stimulus in the preceding task is unrelated to the recognition probe—for example, participants first solve a math problem and then decide whether an unrelated word appeared in the study list. Conversely, in the related paradigm, the stimulus in the preceding task is identical to the recognition probe—for example, participants first solve an anagram and then decide whether the anagram solution appeared in the study list. Verde and Rotello (2004) found that the preceding task induces a criterion shift in both paradigms. However, the participants' sensitivity decreases following the preceding task only in the related paradigm. Thus, the two paradigms have different effects on recognition memory. Previous studies that investigated the effect of task difficulty on the revelation effect used both paradigms (Bornstein & Neely, 2001; Niewiadomski & Hockley, 2001; Verde & Rotello, 2003; Watkins & Peynircioglu, 1990; Westerman & Greene, 1998). As we discussed earlier, these studies either failed to find a link between task difficulty and the revelation effect or confounded task difficulty and task presence. Thus it seems unlikely that the experimental paradigm has much bearing on the present research. However, the present experiments exclusively represent the unrelated paradigm. Consequently, we cannot rule out that the present experiments would have turned out differently, had we used the related paradigm.

Another limitation of the present study concerns the lack of statistically significant effects on the level of individual experiments. The moderating effect of task difficulty on the size of the revelation effect failed conventional levels of significance despite typical or larger than typical sample sizes for this type of experiment. This could be taken as evidence that task difficulty has no influence on the revelation effect. However, this interpretation neglects several points. Hard preceding tasks consistently produced larger revelation effects than easy preceding tasks with the single exception of the criterion estimates in Experiment 4. Further, combining the results of even two or three experiments can drastically decrease the estimation error of effect sizes as compared to individual experiments (Cumming, 2012). When we combined the results of the present experiments, we found a significantly larger revelation effect for hard than for easy preceding tasks. We also found that easy preceding tasks produced a small but reliable revelation effect. Overall, the present experiments are difficult to reconcile with the assumption that the revelation effect is independent of task difficulty. A potential direction for future research is to include more than two levels of task difficulty in a single experiment. Such a study could provide a more precise picture of the link between task difficulty and the revelation effect.

The present experiments also provide a cautionary tale regarding several general methodological issues. The results of the present experiments demonstrate the degree of uncertainty associated with the results of single experiments with samples sizes that are arguably very common in Cognitive Psychology. In the present study, Experiments 2, 4, and 6 are direct replications of Experiments 1, 3, and 5. The results between the original experiments and the replications vary, sometimes considerably (Cumming, 2012, 2014). This issue is not restricted to the present set of experiments but has been highly prevalent in psychological research (Open Science Collaboration, 2015).

The high variability of effect sizes across experiments can have serious consequences. Consider, for example, a scenario in which a set of similar experiments with small variations in the experimental procedure cause p values for the critical analysis on either side of the significance criterion ($\alpha = .05$). Because the significance test suggests that a result is either significant or not, the experiments are considered to be inconclusive (Loftus, 1996). Even worse, small variations in the experimental procedure invite post-hoc hypotheses about what aspect of the procedure made the effect appear in some experiments but not in other experiments. However, a

variation in effect sizes is to be expected even in well executed replications of an experiment (Cumming, 2012, 2014). Another methodological issue concerns the practice of "cherry picking" results for publication, creating publication bias. On the level of individual experiments, the present experiments, when published selectively, could have supported opposing hypotheses: The revelation effect depends versus does not depend on task difficulty. However, when combining all available data, only one conclusion is possible: Hard preceding tasks produce a larger revelation effect than easy tasks.

With regard to revelation effect hypotheses, the present experiments allow several conclusions. According to the global-matching hypothesis the revelation effect should not have occurred in Experiments 3–6 because the arrow-key task has no obvious similarity to the word material in the study list. However, a robust revelation effect appeared in these experiments. Furthermore, the criterion-flux hypothesis suggests that the task difficulty threshold for a revelation effect is low and once this threshold is crossed, the revelation effect appears at uniform strength. This assumption is incompatible with the present experiments. Specifically, when we combined the data of our experiments, a small but reliable revelation effect in the easy condition appeared. Further, this effect was larger in the hard than in the easy condition, which suggests that the revelation effect is gradual rather than all-or-nothing in nature.

Both the familiarity-decrement and discrepancy-attribution hypotheses predicted the data pattern in Experiments 1–6. However, these hypotheses are vague in their description of how task difficulty affects the size of the revelation effect. The familiarity-decrement hypothesis postulates that the preceding task activates memory traces. This activation is assumed to compete with the recognition probe, thus increasing the difficulty of the recognition judgment. However, in the present experiments, the preceding task (e.g., arrow keys) was unrelated to the recognition probe (i.e., a word). Thus, if one assumes that only highly associated memory traces can compete with each other, we should not have observed a revelation effect in the present experiments. Unfortunately, Hicks and Marsh (1998) did not specify which memory traces can compete. Consequently, the present experiments do not necessarily contradict the familiarity-decrement hypothesis but the vague definition of "competition" is in itself problematic.

At first glance the concept of fluency—the ease and speed of information processing—in the discrepancy-attribution hypothesis is more specific about how difficulty interacts with the revelation effect. However, properly testing this hypothesis would require a measure of the temporal development of fluency. Importantly, this measurement would have to be subtle enough so it would not serve as a preceding task on its own. Without such a measure, one can only conclude from the presence of a revelation effect that fluency must have been discrepant enough. However, this argument amounts to circular reasoning, where discrepancy causes the revelation effect, which in turn indicates the occurrence of discrepancy. It can also be difficult to derive predictions from fluency-based accounts due to several mechanisms that have been suggested to govern the attribution of fluency. Although fluency is thought to inform judgments of familiarity, Whittlesea and Williams (2001) suggested that this is only the case for surprising fluency. Similarly, Oppenheimer (2006) suggested that fluency can be discounted if participants become aware of the source of fluency and decide that this source is not relevant for their judgment. Unfortunately, participants are also thought to overdiscount at times; that is, fluency may actually decrease the perceived familiarity of a stimulus. A fluency framework that includes all these concepts is in danger of "explaining" everything but predicting nothing. In addition, the familiarity-decrement and discrepancy-attribution hypotheses fail to account for the results of other studies. Both hypotheses struggle, for example, to explain why the revelation effect reverses in some circumstances, resulting in more "new" responses in the task condition than in the no-task condition (Aßfalg & Bernstein, 2012).

Conclusion

The present experiments support the familiarity-decrement and discrepancy-attribution hypotheses. Conversely, the present experiments do not support other revelation effect hypotheses that have predicted the absence of a revelation effect or the absence of a moderating effect of task difficulty on the size of the revelation effect. A major challenge for future research will be to introduce formalized hypotheses for the cause of the revelation effect that can provide more precise predictions and better testability. Ideally, such hypotheses should be part of existing formal models of memory.

Author note This research was funded by a German Research Foundation grant to A.A. (AS 427/1-1) and by the Canada Research Chairs program to D.M.B. (950-228407).

Appendix

Table 4 Mean (*SD*) hits and false alarms as a function of task difficulty (hard vs. easy) and task presence (task vs. no task) in Experiments 1–6

Experiment/Difficulty		Hits			False Alarms		
		Task	No Task	*d* [95% CI]	Task	No Task	*d* [95% CI]
1	Hard	.68 (.18)	.61 (.21)	0.34* [0.08, 0.61]	.49 (.21)	.41 (.23)	0.35* [0.11, 0.60]
	Easy	.63 (.18)	.64 (.19)	–0.02 [–0.30, 0.27]	.41 (.25)	.35 (.22)	0.25* [0.01, 0.49]
2	Hard	.70 (.18)	.61 (.20)	0.43* [0.07, 0.80]	.38 (.24)	.32 (.25)	0.36* [0.12, 0.59]
	Easy	.67 (.19)	.64 (.18)	0.17 [–0.08, 0.42]	.36 (.21)	.30 (.23)	0.25 [–0.02, 0.52]
3	Hard	.67 (.19)	.63 (.17)	0.24 [–0.03, 0.51]	.43 (.22)	.32 (.21)	0.48* [0.12, 0.83]
	Easy	.71 (.17)	.70 (.19)	0.06 [–0.20, 0.33]	.35 (.20)	.33 (.21)	0.12 [–0.11, 0.35]
4	Hard	.75 (.19)	.72 (.16)	0.12 [–0.17, 0.42]	.39 (.24)	.31 (.22)	0.37* [0.20, 0.54]
	Easy	.73 (.18)	.68 (.21)	0.24 [–0.02, 0.50]	.36 (.22)	.31 (.22)	0.22* [0.00, 0.44]
5	Hard	.73 (.17)	.66 (.17)	0.44* [0.13, 0.74]	.35 (.17)	.30 (.17)	0.33* [0.09, 0.57]
	Easy	.68 (.22)	.70 (.19)	–0.11 [–0.35, 0.14]	.26 (.19)	.28 (.22)	–0.10 [–0.32, 0.12]
6	Hard	.72 (.16)	.67 (.23)	0.24 [–0.01, 0.50]	.49 (.22)	.39 (.21)	0.49* [0.20, 0.78]
	Easy	.59 (.23)	.60 (.19)	–0.01 [–0.29, 0.27]	.41 (.23)	.36 (.25)	0.22 [–0.08, 0.51]

Effect sizes (Cohen's *d*) refer to the size of the revelation effect (i.e., comparison of task vs. no task) separately for hits and false alarms Note that hits and false alarms were generated by dichotomizing recognition confidence judgments (1 = *sure new* to 6 = *sure old*) such that Judgments 4–6 were coded as an "old" response, and Judgments 1–3 were coded as a "new" response; * $p < .05$

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

Aßfalg, A. (2017). Revelation effect. In R. Pohl (Ed.), *Cognitive illusions: Intriguing phenomena in judgment, thinking, and memory* (pp. 339–356). Abingdon, UK: Routledge.

Aßfalg, A., & Bernstein, D. M. (2012). Puzzles produce strangers: A puzzling result for revelation-effect theories. *Journal of Memory and Language, 67,* 86–92. doi:10.1016/j.jml.2011.12.011.

Aßfalg, A., & Nadarevic, L. (2015). A word of warning: Instructions and feedback cannot prevent the revelation effect. *Consciousness and Cognition, 34,* 75–86. doi:10.1016/j.concog.2015.03.016.

Bernstein, D. M., Whittlesea, B. W. A., & Loftus, E. F. (2002). Increasing confidence in remote autobiographical memory and general knowledge: Extensions of the revelation effect. *Memory & Cognition, 30,* 432–438. doi:10.3758/BF03194943

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.

Bornstein, B. H., & Neely, C. B. (2001). The revelation effect in frequency judgment. *Memory & Cognition, 29,* 209–213.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3,* 37–60. doi:10.3758/BF03210740

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112,* 155–159. doi:10.1037/0033-2909.112.1.155

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Routledge.

Cumming, G. (2014). The new statistics why and how. *Psychological Science, 25,* 7–29. doi:10.1177/0956797613504966

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41,* 1149–1160. doi:10.3758/BRM.41.4.1149

Hicks, J. L., & Marsh, R. L. (1998). A decrement-to-familiarity interpretation of the revelation effect from forced-choice tests of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1105–1120. doi:10.1037/0278-7393.24.5.1105

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 302–313. doi:10.1037/0278-7393.21.2.302

Hockley, W. E., & Niewiadomski, M. W. (2001). Interrupting recognition memory: Tests of a criterion-change account of the revelation effect. *Memory & Cognition, 29,* 1176–1184.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340. doi:10.1037/0096-3445.110.3.306

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Loftus, G. R. (1996). Psychology will be a much better science when we change the way we analyze data. *Current Directions in Psychological Science, 5,* 161–171. doi:10.1111/1467-8721.ep11512376

Macmillan, N. A., & Creelman, C. (2005). *Detection theory: A user's guide*. Mahwah, NJ: Erlbaum.

Niewiadomski, M. W., & Hockley, W. E. (2001). Interrupting recognition memory: Tests of familiarity-based accounts of the revelation effect. *Memory & Cognition, 29,* 1130–1138.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349,* 943. doi:10.1126/science.aac4716

Oppenheimer, D. M. (2006). Consequences of erudite vernacular utilized irrespective of necessity: Problems with using long words needlessly. *Applied Cognitive Psychology, 20,* 139–156. doi:10.1002/acp.1178.

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

Peynircioğlu, Z. F., & Tekcan, A. I. (1993). Revelation effect: Effort or priming does not create the sense of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 382–388. doi:10.1037/0278-7393.19.2.382

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review, 8,* 364–382. doi:10.1207/s15327957pspr0804_3

Rhodes, M. G., & Jacoby, L. L. (2007). On the dynamic nature of response criterion in recognition memory: Effects of base rate, awareness, and feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 305–320. doi:10.1037/0278-7393.33.2.305.

Singer, M., & Wixted, J. T. (2006). Effect of delay on recognition decisions: Evidence for a criterion shift. *Memory & Cognition, 34,* 125–137. doi:10.3758/BF03193392

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50. doi:10.1037/0096-3445.117.1.34

Verde, M. F., & Rotello, C. M. (2003). Does familiarity change in the revelation effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 739–746. doi:10.1037/0278-7393.29.5.739

Verde, M. F., & Rotello, C. M. (2004). ROC curves show that the revelation effect is not a single phenomenon. *Psychonomic Bulletin & Review, 11,* 560–566.

Watkins, M. J., & Peynircioglu, Z. F. (1990). The revelation effect: When disguising test items induces recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1012–1020. doi:10.1037/0278-7393.16.6.1012

Westerman, D. L., & Greene, R. L. (1996). On the generality of the revelation effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1147–1153. doi:10.1037/0278-7393.22.5.1147.

Westerman, D. L., & Greene, R. L. (1998). The revelation that the revelation effect is not due to revelation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 377–386. doi:10.1037/0278-7393.24.2.377.

Whittlesea, B. W. A., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language, 29,* 716–732. doi:10.1016/0749-596X(90)90045-2

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? A discrepancy-attribution account of feelings of familiarity. *Acta Psychologica, 98,* 141–165. doi:10.1016/S0001-6918(97)00040-1.

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

Whittlesea, B. W. A., & Williams, L. D. (2001). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 3–13. doi:10.1037/0278-7393.27.1.3

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers, 20,* 6–10. doi:10.3758/BF03202594