

# Can pupillometry distinguish accurate from inaccurate familiarity?

Carlos A. Gomes<sup>1,2</sup>  | Daniela Montaldi<sup>1</sup> | Andrew Mayes<sup>1</sup>

<sup>1</sup>School of Biological Sciences, Division of Neuroscience & Experimental Psychology, University of Manchester, Manchester, UK

<sup>2</sup>Department of Neuropsychology, Ruhr-University Bochum, Bochum, Germany

## Correspondence

Andrew Mayes, School of Biological Sciences, Division of Neuroscience & Experimental Psychology, University of Manchester, M13 9PL Manchester, UK.  
Email: andrew.mayes@manchester.ac.uk

## Funding information

This work was supported by the Fundação para a Ciência e a Tecnologia under Grant SFRH/BD/41637/2007 and the University of Manchester

## Abstract

Pupillometry, the measurement of pupil diameter, has become increasingly popular as a tool to investigate human memory. It has long been accepted that the pupil is able to distinguish familiar from completely novel items, a phenomenon known as “pupil old/new effect”. Surprisingly, most pupillometric studies on the pupil old/new effect tend to disregard the possibility that the pupillary response to familiarity memory may not be entirely exclusive. Here, we investigated whether the pupillary response to old items correctly judged familiar (hits; accurate familiarity) can be differentiated from the pupillary response to new items wrongly judged familiar (false alarms; inaccurate familiarity). We found no evidence that the two processes could be isolated, as both accurate and inaccurate familiarity showed nearly identical mean and across-time pupillary responses. However, both familiarity hits and false alarms showed pupillary responses unequivocally distinct from those observed during either recollection or novelty detection, which suggests that the pupil measure of familiarity hits and/or false alarms was sufficiently sensitive. The pupillary response to false alarms may have been partially driven by perceptual fluency, since novel objects incorrectly judged to be old (i.e., false alarms) showed a higher degree of similarity to studied images than items correctly judged as novel (i.e., correct rejections). Thus, our results suggest that pupil dilation may not be able to distinguish accurate from inaccurate familiarity using standard recognition memory paradigms, and they also suggest that the pupillary response during familiarity feelings may also partly reflect perceptual fluency.

## KEYWORDS

false memory, familiarity, perceptual fluency, pupil dilation, recognition memory

## 1 | INTRODUCTION

The use of pupillometry (i.e., the measurement of pupil diameter) in the investigation of human memory has witnessed a growing interest over the past two decades. Several pupillometric studies have shown that the pupil can discriminate between studied and unstudied stimuli during

recognition memory tests (Heaver & Hutton, 2011; Kafkas & Montaldi, 2011; Otero et al., 2011; Võ et al., 2008), even in situations where individuals are not able to distinguish old from new stimuli (Gomes et al., 2015; Laeng et al., 2007). Pupillometry has also been useful in helping dissociate familiarity from recollection in recognition tests (Kafkas & Montaldi, 2012; Otero et al., 2011), two kinds of recognition

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Psychophysiology published by Wiley Periodicals LLC on behalf of Society for Psychophysiological Research.



memory that are characterized by memory for the item alone without its study context and recall memory for the item in its study context, respectively.

Despite copious evidence suggesting a strong link between accurate familiarity (indexed by familiarity hits) and increases in pupil diameter, much less is known about the relationship between accurate familiarity and inaccurate familiarity (indexed by false alarms). Some evidence suggests that recognition false alarms can be induced by enhancing processing fluency (e.g., via masked priming) of items that have not been studied in an earlier phase of an experiment. The unexpected increase in fluency causes participants to endorse those fluent items as old, regardless of their true study status (Jacoby & Whitehouse, 1989; Johnston et al., 1985; Whittlesea, 1993; Whittlesea & Williams, 2000).

Neuroimaging research indicates that accurate familiarity and fluency-driven inaccurate familiarity memory may recruit similar neural mechanisms. Using fMRI, Dew and Cabeza (2013) found that reduced activity in the perirhinal cortex, which, for decades, had been considered a hallmark of familiarity memory, was also observed for unstudied words that were primed during the test phase. These and several other findings (Gomes et al., 2016, 2017, 2019; Wang et al., 2016) suggest that the neural substrates of accurate and fluency-based inaccurate recognition memory are likely mediated by a common neural network. Indeed, computational models have indicated that, if certain assumptions hold, a single-memory system is capable of supporting both recognition memory and fluency (Berry et al., 2008, 2012).

If enhanced fluency can induce false memories, what effect has increased processing fluency on pupillary responses? We are only aware of one study (Kuchinke et al., 2009) that examined the relationship between processing fluency and pupil dilation. In this study, participants were shown Cubist paintings which varied in content accessibility (low, medium and high), and were asked to press a button as soon as they recognized a concrete object in the painting (e.g., the face of a woman). The results showed that the fluently-processed paintings were responded to faster, and this effect was accompanied by larger pupil dilation at the point of object recognition. Thus, this study suggests that enhancing processing fluency can lead to increased pupil dilation. The question remains, however, whether the enhanced fluency that triggers a false alarm during a recognition memory test generates a pupillary response that resembles that of a typical familiarity-based pupillary response.

The idea that accurate and inaccurate familiarity could potentially exhibit similar pupillometric behavior has rarely been considered. There exist, however, a few studies which reported larger pupil dilation for new stimuli judged “old” (i.e., false alarms) than new stimuli judged “new” (i.e., correct rejections) (Brocher & Graf, 2017; Montefinese et al., 2013; Otero et al., 2011). In addition, these studies also found that false alarms showed less dilation than correct recognition memory

(truly old items correctly categorized as old). However, participants only had the option to judge an item as either “old” or “new,” rather than recollected, familiar or new/unstudied, so pupil dilation for “old” judgments might have reflected the combined contribution of recollection and familiarity memory.

Worryingly, most pupillometric studies on human memory often interpret pupil dilation differences between familiarity hits and correctly-rejected responses as evidence that it *specifically* indicates accurate familiarity. However, if false alarms produce a similar pupillary response to that observed with familiarity hits, then the specificity of pupil dilation to familiarity memory (and perhaps recollection) may need revisiting. Indeed, in a recent study from Mill et al. (2016), participants were given highly predictive cues of the true old/new status of words before making their recognition memory judgments. They found that when participants were specifically cued that the status of the word would be “new”, the pupil showed a stronger dilation for “old” than “new” responses, and this was true for both correct (hits > correct rejections) and incorrect (false alarms > misses) judgments. In contrast, for “likely old” cues, no difference between “old” and “new” responses was found. The authors concluded that pupil old/new effects do not signal accurate recognition, but, instead, could reflect participants’ subjective sense of recognition.

It is, therefore, critical to understand how specific pupil dilation that accompanies accurate familiarity really is, by comparing the pupillometric behavior of familiarity hits with familiarity-based false alarms. In a previous study, we provided evidence that implicit memory can be dissociated from accurate familiarity at the pupil level (Gomes et al., 2015). In that study, we showed that even though forgotten items (i.e., misses) displayed larger pupils than correctly rejected new items (i.e., correct rejections), this pupil dilation was reduced compared to that observed for familiarity hits (even when appropriately matched for difficulty and familiarity strength). Here, we re-analyzed the same dataset but examined the relationship between familiarity hits (accurate familiarity) and familiarity-based false alarms (inaccurate familiarity) instead. Our familiarity-only procedure focused attention on making familiarity decisions about stimuli studied under incidental conditions, likely reducing recollection memory so that relatively few stimuli were recollected (Mayes et al., 2007; Montaldi et al., 2006). This procedure may have also increased the degree to which fluency processing contributed to familiarity decisions.

If fluency modulates pupil dilation as some research suggests (Kuchinke et al., 2009), then it is plausible to assume that pupillary responses to both familiarity hits and false alarms are equally affected by fluency, and they may show similar pupillary behavior. However, if fluency plays a minimal role, then pupil dilation for familiarity hits and false alarms may well be different. Alternatively, if the brain fails to distinguish accurately whether an item is old or new, false alarms may still produce a brain response that

translates into indistinguishable pupillometric behavior from familiarity hits, regardless of how fluently the items were processed.

## 2 | METHOD

### 2.1 | Participants, materials, and procedure

Power analysis revealed that at least 45 participants would be required to achieve a power of .80. For this power calculation, we used the effect size of the difference in the pupillary time courses between familiarity hits and misses in Gomes et al. (2015). In the present study, 90 participants were recruited. Twelve participants had to be excluded due to either technical difficulties in the discrimination of the pupil or exaggerated movement during the tasks.

We followed the recommendations given by Goldinger and Papesch (2012) when applying pupillometry to the study of human memory, which included the usage of: (1) colorless, low-contrast stimuli, (2) relatively long stimulus exposure and interstimulus interval, (3) baseline-corrected diameter analyses, and (4) complementary pupillometric metrics. We have described the materials and procedure in detail in Gomes et al. (2015). In short, participants were presented with either 50 (Experiment 1) or 100 (Experiments 2 and 3) object pictures at study, and either counted the number of red dots (either one or two) that were flashed within the image borders (Experiment 1 and 2) or decided if the object was bigger or smaller than a shoebox (Experiment 3). At test, participants saw the studied pictures randomly interspersed with 100 novel pictures (50 in Experiment 1). They were instructed to perform a recognition memory task, in which they decided whether each picture was recollected, familiar or new. During the test phase, a fixation cross was presented for 3,000 ms, followed by a blank screen for 500 ms (baseline) and finally the picture of an object, which remained on the screen until participants responded. Pupil data were recorded for each individual trial, starting at baseline onset until participants' response.

### 2.2 | Data cleaning, reduction, and analysis

Blinks and other artifacts were removed from each pupillary trace and linear interpolation replaced data points for a particular trace that deviated from the mean by more than plus/minus 2 standard deviations. An unweighted 5-point moving average filter was applied to the data. Furthermore, trials with reaction times (RTs) that were greater or smaller than 2 standard deviations were also excluded from the analysis.

In all our analyses, the mean pupillary response was subtracted from the averaged baseline period (500 ms prior

to stimulus onset). These baseline-corrected pupillary responses were calculated for recollection hits (Rs), familiarity hits (Fs), familiarity false alarms (FAs), correct rejections (CRs), and misses (Ms). In order to reduce the effects of conscious effort and familiarity strength on pupil dilation, we applied an RT-matching procedure whenever any two conditions were contrasted (see Gomes et al., 2015, for details). This procedure consisted of iterating over every trial of a particular category (e.g., FA) in order to find a trial of a different category (e.g., CR) with a similar RT. Positive differences were balanced against negative differences, such that the net mean RT for the categories being matched would be identical (Gomes et al., 2015; Spencer et al., 2009). Only participants with at least 3 valid trials in any given response category were considered.

Note that the significance, as well as direction of the results, were unchanged when the unmatched data were analyzed instead (see Supplementary Information). However, given that RTs for some of the categories were significantly different from each other (e.g., FAs vs. CRs), we decided to present the matched data in the main text.

There were no effects interacting with the experiment factor in any of the analyses performed (all  $F$ s < 1.90,  $p$ s > .16). Thus, to simplify the analyses and leverage the larger sample size, we collapsed the data across the experiment factor.

### 2.3 | BOLAR index of image similarity

For the analysis of image similarity, we computed the bank of local analyzer responses (BOLAR; Zelinsky, 2003).<sup>1</sup> The BOLAR model isolates the visual similarity component by decomposing a stimulus into featural visual dimensions such as color, orientation, and spatial scale. This is achieved by the use of 108 multiscale Gaussian-derivative filters (GDFs) that code the visual features of the image. Large-scale filters extract low spatial frequency patterns in an image (e.g., overall shape of an object), whereas small-scale filters extract high spatial frequency patterns (e.g., fine details of an object). These 36 GDFs are then applied to the three color/intensity channels ( $36 * 3 = 108$  filters), thereby, capturing chromatic information. Because this is done across all pixels in an image, implementing this procedure results in a BOLAR vector for each individual pixel. Once the BOLAR vectors of any two images are known,

<sup>1</sup>Note that we did not employ this algorithm in our previous study (Gomes et al., 2015), because the focus of that study was on long-term priming rather than perceptual fluency. In the present report, however, we examined the possibility that perceptual fluency affected the pupillary responses of FAs. Therefore, we used the BOLAR index here to measure perceptual similarity between images.

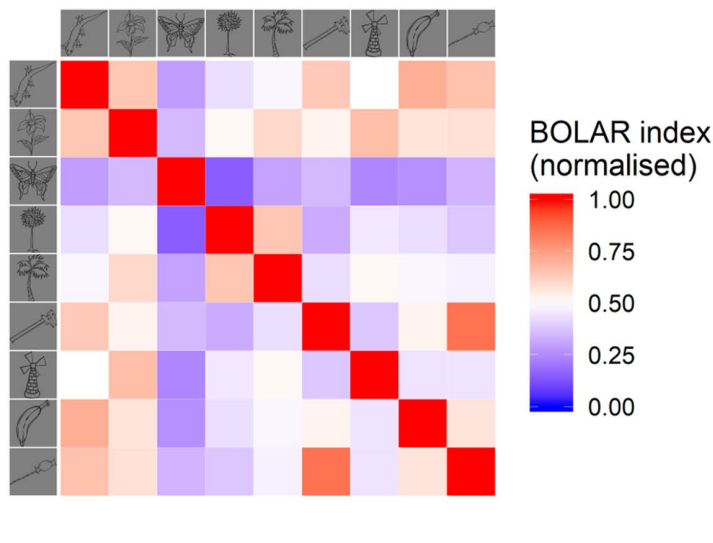
the normalized Euclidean distance ( $E$ ) between those BOLAR vectors can be computed, giving a map of image differences. This difference map is then summarized by squaring and summing each signal, and subsequently taking the square root of that value (Equation 1a). The resulting scores are then normalized in order to bring all values to the range  $[0, 1]$ , and then subtracted by 1 (such that low values indicated low similarity and high values indicated high similarity; Equation 1b).

$$E = \frac{1}{n} \sqrt{\sum_{k=1}^n [BV_k - BV'_k]^2} \quad (1a)$$

$$BI_i = 1 - \left( \frac{x_i - \min(x)}{\max(x) - \min(x)} \right) \quad (1b)$$

Equation 1. (a) Calculation of the Euclidean distance ( $E$ ) between any two images (Zelinsky, 2003).  $n$  represents the dimensionality of the BOLAR representation;  $BV$  and  $BV'$  are the two BOLAR vectors being compared. (b) The  $E$ s for each paired comparison ( $x_i$ ) were normalized within their own response category  $x$  (e.g., FA), so that values had a score between 0 and 1, with scores closer to 1 representing higher similarity—this is the Bolar index (BI).

Figure 1 (left) shows an example of the results of the BOLAR analysis for a subset of images with varying degrees of similarity of perceptual features.

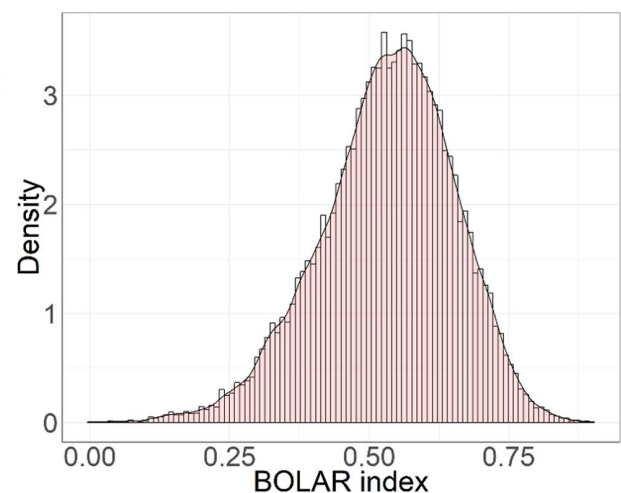


### 3 | RESULTS

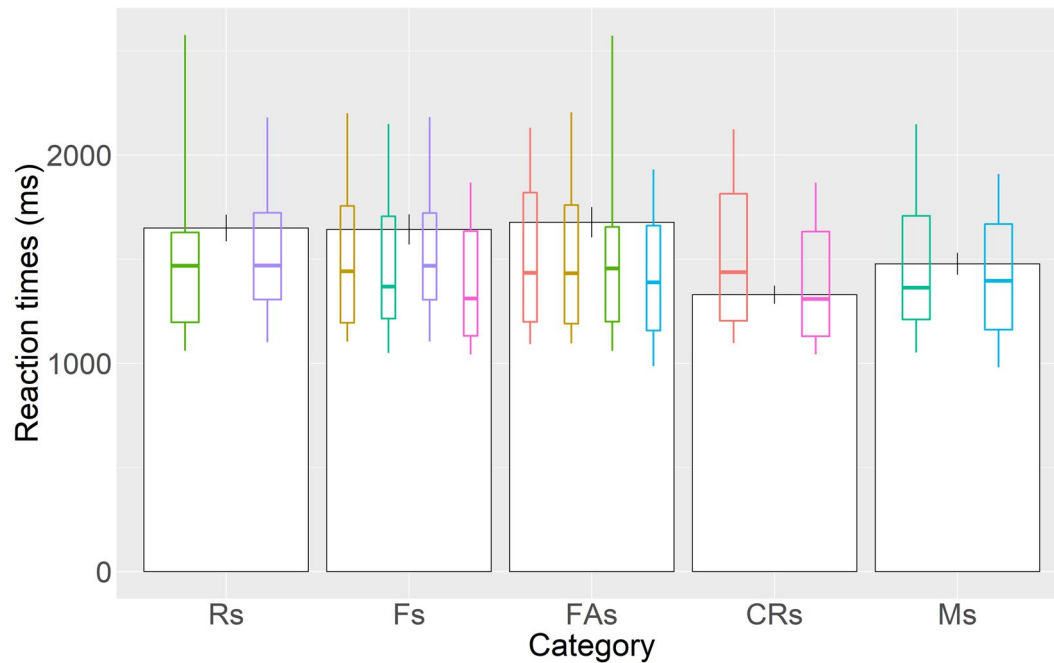
Familiarity memory performance, defined as  $[(Fs - FAs)/(Fs + FAs)]$ , was well above chance for both the collapsed data (mean = .55,  $SD = .24$ ;  $t(77) = 20.18$ ,  $p < .001$ ,  $d = 2.31$ ) and the individual experiments (see Gomes et al., 2015).

After RT-matching (see methods and Gomes et al., 2015), RTs were not significantly different among the different response categories (see Figure 2).

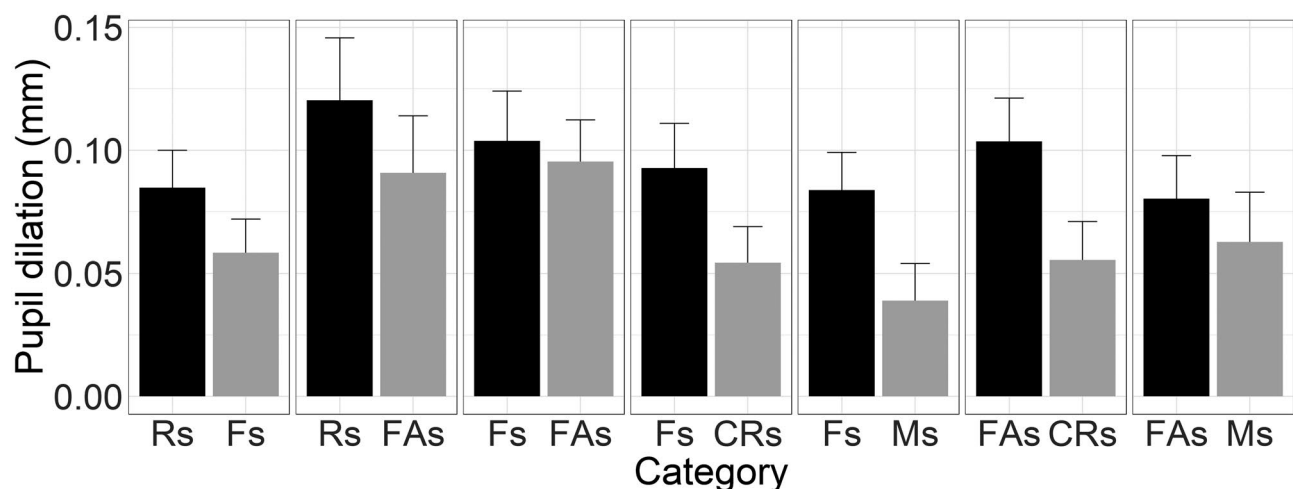
Our main interest in the present study was to determine whether inaccurate familiarity (as indexed by FAs), could be distinguished from accurate familiarity (as indexed by Fs). For that purpose, we initially compared Fs with properly RT-matched FAs. There was no difference between the two categories with respect to mean pupil diameter,  $F(1,70) = .41$ ,  $p = .68$  (see Figure 3). Even though we did not observe significant interactions with the Experiment factor (see Methods section), we decided to compare Experiments 1 and 2 with Experiment 3, given that they differed in the encoding task: Experiments 1 and 2 used a more perceptual orienting task (dot task), whereas Experiment 3 used a more semantic orienting task (size-judgment task). First, as with the collapsed data, we did not find a statistically significant difference between Fs and FAs in either experiment ( $ts < 1.00$ ,  $ps > .33$ ,  $ds < .21$ ). In addition, there was no pupillary response difference for the Fs versus FAs comparison between Experiment 1/2 and Experiment 3 ( $t(69) = 1.11$ ,  $p > .27$ ,  $d = .29$ ). This was true even though memory performance was substantially better for the size-judgment than the dot task, and RTs for



**FIGURE 1** Left: Graphical depiction of the BOLAR index as a measure of the similarity between pairs of pictures. The same nine exemplar items were selected and arranged along the vertical and horizontal axis of a 9×9 matrix. Colors indicate different BOLAR scores, in that larger values (red) indicate higher similarity. The diagonal of the matrix has a value of 1 since the images being compared are identical. Note how BOLAR accurately captures similarity based on the overall shape (e.g., the shape of the pen and the thermometer), as well as finer details (e.g., the petals of the flower and the windmill sails). Right: Density distribution of BOLAR scores for all possible pairwise comparison of two images



**FIGURE 2** Mean raw reaction times for recollection hits (Rs), familiarity hits (Fs), false alarms (FAs), and correct rejections (CRs). Color boxplots represent mean reaction times after implementing the matching procedure (see Methods and Gomes et al. (2015) for details). Each color is associated with a specific matching between two categories (e.g., purple boxplots represent the matched reaction times of Rs and Fs, red boxplots represent the matched reaction times of FAs and CRs, etc.). Error bars represent the standard error of the mean. Rs = Recollection hits, Fs = Familiarity hits, FAs = False alarms, CRs = Correct rejections, Ms = Misses

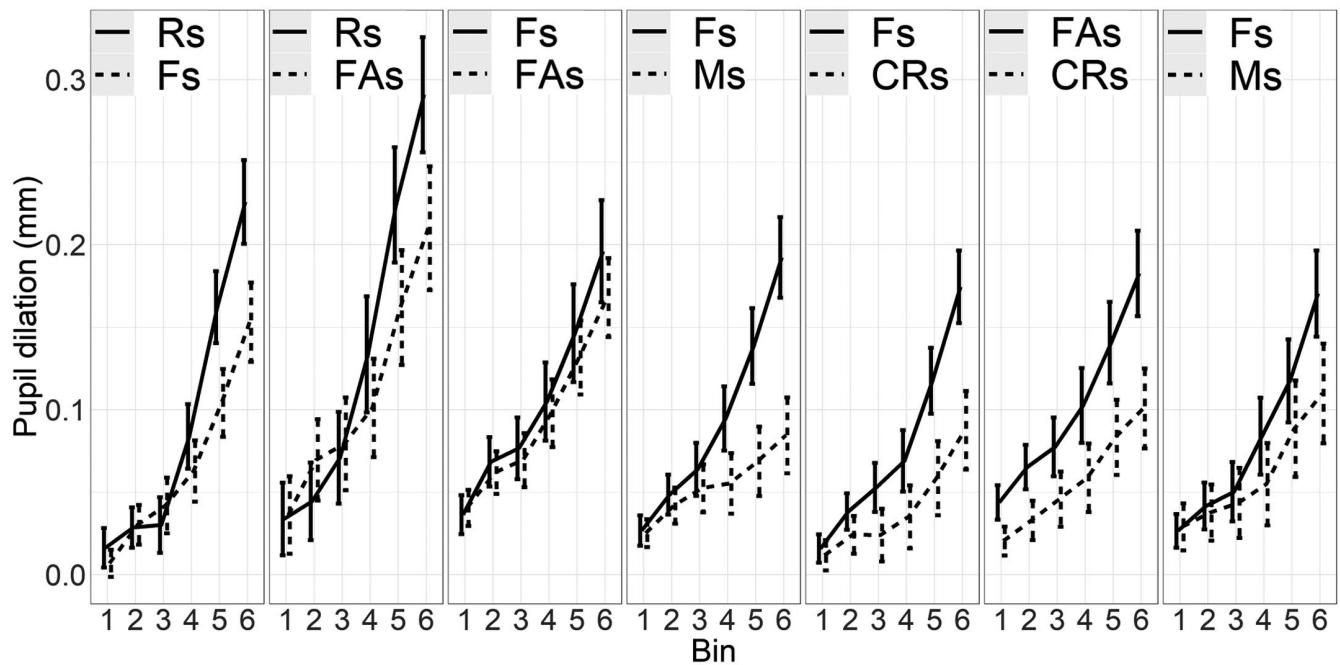


**FIGURE 3** Mean baseline-corrected pupil dilation for RT-matched categories. The categories are RT-matched within each facet (e.g., for the first facet, Rs are RT-matched with Fs). Error bars represent the standard error of the mean. Rs = Recollection hits, Fs = Familiarity hits, FAs = False alarms, CRs = Correct rejections, Ms = Misses

both Fs and FAs were also significantly faster for the size-judgment than the dot task (all  $t$ s  $> 2.94$ ,  $p$ s  $< .005$ ). This result is interesting because it suggests that the pupillary response between Fs and FAs does not differ, even under experimental conditions where F memory is substantially stronger. Therefore, we continued data analyses with the collapsed data only.

Bayesian statistics allow the assessment of how much support we have for the null hypothesis in comparison to the alternative hypothesis. The Bayes Factor (BF), in particular, represents the evidence in the data favoring one hypothesis over another. We were interested in comparing the hypothesis of larger pupil dilation for Fs than FAs against the null hypothesis of no difference between the two categories.





**FIGURE 4** Mean baseline-corrected pupil dilation for RT-matched categories. The categories are RT-matched within each facet (e.g., for the first facet, Rs are RT-matched with Fs). Each bin represents a standardized time window, ranging from stimulus onset (Bin 1) until participants' response (Bin 6; see text for further details). Note that the first time bin corresponds to the average of the first few milliseconds of pupillary recordings. Because the pupil dilates above baseline in most trials, the first bin reflects this averaged initial dilation after stimulus onset, which is the reason why the waveforms do not start at exactly 0. Error bars represent the standard error of the mean. Rs = Recollection hits, Fs = Familiarity hits, FAs = False alarms, CRs = Correct rejections, Ms = Misses.

Conventionally, BF values equal or larger than 3 or equal or smaller than  $1/3$  represent substantial evidence, and anything between  $1/3$  and 3 are considered inconclusive evidence. In our case, the BF was .3, which would be interpreted as our data supporting the null hypothesis.

Next, we compared the pupillary time courses of Fs versus FAs. For this analysis, the pupil data for each individual trial was split into six bins, each containing an equivalent number of pupil recordings.<sup>2</sup> This procedure allowed us to standardize different trials that had different RTs (all three experiments were self-paced), and, thus, contained different numbers of pupil recordings. Each bin was then averaged across all trials for a particular category, thus providing a binarized (averaged) time course for each participant and response category (Gomes et al., 2015).

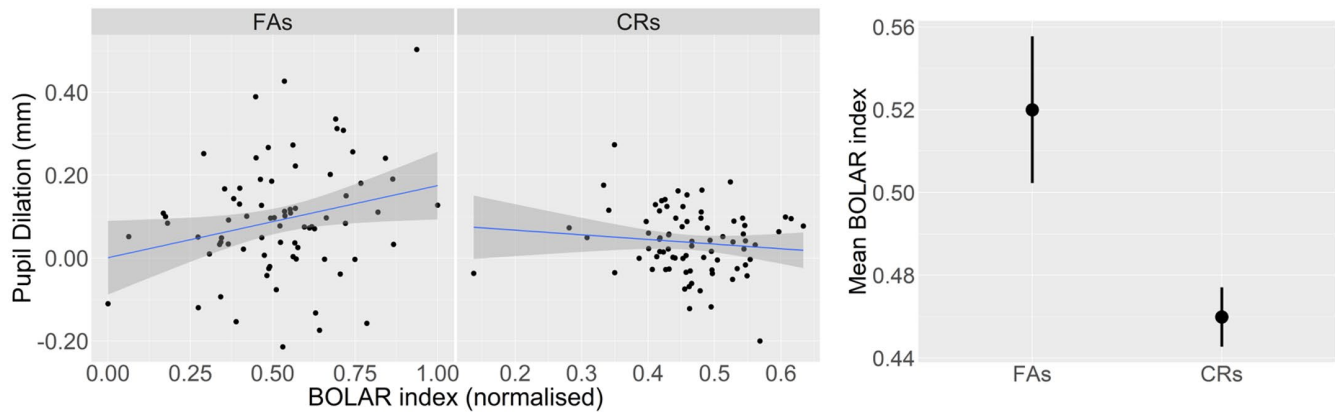
The time course analysis only revealed a significant main effect of Bin,  $F(1.44, 102.04) = 27.75$ ,  $p < .001$ ,  $\eta^2_p = 1$ , but no critical Bin  $\times$  Category interaction,  $F(1.62, 114.86) = 1.43$ ,  $p = .24$ ,  $\eta^2_p = .27$  (see Figure 4). Unsurprisingly, the subsequent post-hoc analysis did not reveal any significant differences between the two categories at any bin level (all  $t$ s  $< 1.25$ ,  $p$ s  $> .21$ ). The BF analysis also found strong evidence favoring the model

with the main effect of bin relative to any other model. In fact, the reduced model with only the main effect of Bin was favored against the full model (which included the Bin  $\times$  Category interaction) by a factor of more than 200 ( $BF_{\text{main}}/BF_{\text{interaction}} = 245.04 [\pm 4.76\%]$ ).

In order to demonstrate that the lack of a pupillary response difference between Fs and FAs was not the result of insensitivity of either Fs or FAs, we compared these two categories to RT-matched CRs. Both Fs and FAs differed statistically from their corresponding matched CRs ( $F$ s  $> 5.81$ ,  $p$ s  $< .02$ ,  $\eta^2_p > .07$ ). In both analyses there was also a significant main effect of Bin (both  $F$ s  $> 18.49$ ,  $p$ s  $< .001$ ,  $\eta^2_p > .20$ ), as well as a significant Bin  $\times$  Category interaction (both  $F$ s  $> 3.71$ ,  $p$ s  $< .032$ ). Post-hoc tests revealed that both bins 5 and 6 significantly differed between Fs and CRs as well as between FAs and CRs ( $t$ s  $> 2.19$ ,  $p$ s  $< .032$ ). Similarly, the comparison against RT-matched Ms yielded a significant main effect of Bin and a Bin  $\times$  Category interaction for both Fs versus Ms as well as FAs versus Ms (main effect:  $F$ s  $> 14.69$ ,  $p$ s  $< .001$ ,  $\eta^2_p > .17$ ; interaction:  $F$ s  $> 3.69$ ,  $p$ s  $< .03$ ,  $\eta^2_p > .05$ ). Post-hoc tests indicated that bins 5 and 6 significantly differed between Fs and Ms,  $t$ s  $< 3.51$ ,  $p$ s  $< .002$ , whereas only bin 6 differed between FAs and Ms,  $t = 1.96$ ,  $p = .05$ .

The previous analysis revealed that the pupil size for both F and FA responses can be differentiated from the pupil size of CR and M responses, in the sense that the former two

<sup>2</sup>We performed this analysis using different number of bins (up to 10), but the results did not change significantly. Thus, we show here the analysis using six bins to be consistent with Gomes et al. (2015).



**FIGURE 5** Left: Relationship between the BOLAR index (image similarity measure) of FAs and CRs (vs. studied pictures) and pupil dilation. Each point in this figure represents the mean FA BOLAR index (left facet) and mean CR BOLAR index (right facet) for a given participant. Right: Mean BOLAR index separately for FAs and CRs. Shaded area on the left plot and error bars on the right plot represent the standard error of the mean. FAs = False alarms, CRs = Correct rejections

response categories showed larger pupil diameter than the latter two response categories. There is also evidence that F responses show a reduction in pupil diameter relative to R responses (Kafkas & Montaldi, 2012; Otero et al., 2011). Thus, we compared the mean pupillary response between (1) RT-matched Rs and Fs, as well as between (2) RT-matched Rs and FAs. The main effect of category was not significant for either contrast 1 or 2 (both  $F_s < .83$ ,  $p > .37$ ,  $\eta_p^2 < .02$ ). However, for both contrasts, a 2 Category  $\times$  6 Bin of the pupillary time course revealed a significant main effect of Bin,  $F_s > 29.05$ ,  $p_s < .001$ ,  $\eta_p^2 > .38$ , as well as a more interesting Category  $\times$  Bin interaction,  $F_s > 3.56$ ,  $p_s < .03$ ,  $\eta_p^2 > .07$ . As seen in Figure 3, this interaction is mostly the result of Rs showing a noticeably significantly larger increase in pupil dilation than either Fs or FAs in bin 6 ( $t_s > 2.24$ ,  $p_s < .03$ ).

Taken together, the above analyses suggest that FAs behaved much like RT-matched Fs in that (1) these two categories did not differ in terms of either mean pupil size or pupillary time course, and (2) they both differed from their respective RT-matched CRs, Ms, and Rs in a very similar way.

Recent research has shown that fluency-based inaccurate familiarity can trigger a similar brain mechanism to that which produces accurate familiarity (Dew & Cabeza, 2013; Gomes et al., 2019; Voss et al., 2012). One possibility could be that participants experienced enhanced perceptual fluency during the processing of certain novel pictures, which was then used as a cue for distinguishing old from new items (Johnston et al., 1985; Whittlesea, 1993). In order to test if perceptual fluency could partly explain the pupillometric similarity between F and FA responses, we computed the BOLAR index (an indicator of image similarity; see Methods) for every FA and CR image and for all participants. We assumed that the larger the BOLAR index between any given new picture and studied pictures was (i.e., the more similar, on average, a new picture was to studied pictures) the more fluently that new picture would be processed, and that this relationship would only be observed

with FAs. It follows, then, that if an increase in perceptual fluency caused participants to produce FAs, we should observe (1) an association between pupil size and image similarity for FAs but not CRs, and (2) overall larger similarity between studied and FA pictures than studied and CR pictures.

First, we computed a BOLAR index between each individual FA/CR image and each studied image. Then, all BOLAR indices for that FA/CR image were averaged. Therefore, every FA/CR image had an associated BOLAR index, which represented its average Euclidean distance to all studied images. Finally, we calculated the mean of these averaged BOLAR indices for both FAs and CRs, yielding a mean FA BOLAR index and a mean CR BOLAR index for each participant.

We restricted our analysis to CRs and FAs. Correct Fs were excluded because processes other than fluency-based similarity to studied stimuli may well be involved and affect pupil size, and, thus, would introduce a confound. If so, a significant correlation would probably not reflect the effects of fluency alone. In contrast, both CR and FA images were novel, and so, a comparison between these two categories is not biased by prior study status. We initially performed separate correlational analyses in order to determine if mean pupil diameter and the BOLAR index were positively correlated in either response category. As predicted, the relationship between BOLAR index and pupil diameter was significant for FAs,  $r = .25$ ,  $p = .016$ , whereas it was not for CRs,  $r = -.11$ ,  $p = .83$ , suggesting that the relationship between higher similarity and larger pupil dilation was only observed for FAs (see Figure 5, left). The correlation coefficient for FAs was significantly different from the correlation coefficient for CRs,  $z = 2.19$ ,  $p = .01$ .

In addition, we found a significantly larger similarity index for FAs than CRs ( $t = 2.46$ ,  $p = .016$ ,  $d = .28$ ; see Figure 5, right), suggesting that, on average, studied pictures were more similar to FA than CR pictures. We decided to explore this effect further by analyzing how the BOLAR index

**TABLE 1** Results from the multilevel logistic regression model predicting category membership (FA, CR) from pupillary response and BOLAR index predictors (Model 1), and the multilevel regression model predicting pupil dilation from BOLAR index and response category (Model 2)

	<i>T</i>	<i>SE</i>	<i>p</i> -Value
Model 1			
Intercept	−18.75	.10	<.001
Pupil dilation	5.28	.15	<.001
BOLAR index	2.55	.00	.01
Pupil dilation × BOLAR index	.64	.00	.52
Model 2			
Intercept	−1.10	.01	.279
BOLAR index	5.37	.00	.002
Category	3.12	.00	<.001
Pupil dilation × BOLAR index	.20	.00	.84

Abbreviation: *SE*, standard error.

related to making or not making an FA on a trial-by-trial basis. For that aim, we performed a multilevel logistic regression analysis using the lme4 package (Bates et al., 2015) in R (www.r-project.org). We created a model with the response category (CR, FA) as the dependent variable, the BOLAR index for each individual FA/CR image, the mean pupil response, and the interaction between the two as predictors, and participant as a random effect. The idea behind this analysis was to try to predict category membership on the basis of the BOLAR index at each trial. As expected, both mean pupil dilation and BOLAR index were significant predictors of category membership (see Model 1 in Table 1). In other words, the higher the BOLAR index for a given image (i.e., the more similar that image was to studied images), the more likely the model categorized it as an FA. The interaction between pupil dilation and BOLAR index was nonsignificant.

We also built a multilevel regression model with pupil dilation as the dependent variable and response category, BOLAR index and the interaction as predictors, but, similar to Model 1, only the main effects were significant (see Model 2 in Table 1).

Thus, the results of these two analyses suggest that one potential reason that participants made FAs may have been due to the greater similarity that FA pictures had with the study items.

## 4 | DISCUSSION

In this study, we examined the relationship between the pupillary response to familiarity hits (accurate familiarity) and

familiarity false alarms (inaccurate familiarity). Our results unequivocally showed that the mean pupil diameter and pupillary time courses for Fs and RT-matched FAs were indistinguishable from one another. Since FAs correspond to new items that were incorrectly judged to be “old,” and given that both Fs and FAs differed statistically from both RT-matched CRs and Rs, this result suggests that increased pupil dilation during recognition memory tests is not uniquely related to successful recognition memory (see also Mill et al., 2016). Rather, our results imply that accurate and inaccurate familiarity may involve the operation of a common neural system that triggers a pupillary response of the same kind.

Considering that a few studies have found that familiarity hits show different pupillometric behavior from false alarms (Brocher & Graf, 2017; Montefinese et al., 2013; Otero et al., 2011), our finding might seem surprising. In those studies, however, the old/new recognition memory test administered did not strictly encourage the use of familiarity. In fact, it is likely that the pupillary response to recognized items in those studies was the result of the combined contribution of recollection and familiarity memory. Given that recollection produced larger pupil dilation than familiarity, as our and other studies have found (Kafkas & Montaldi, 2011, 2012; Otero et al., 2011), it is not surprising that correct recognition in the above studies was associated with an increased pupillary response relative to inaccurate memory (which may be predominantly based on a sense of familiarity). In contrast, our study minimized the influence of recollection processes during the recognition memory test by using the familiarity-only procedure, which is a modification of the remember/know procedure. Thus, the discrepancies between those and our study could reflect these differences in methodology.

Another study that measured pupil responses and also used the familiarity-only procedure with object stimuli had a different aim that involved comparing oldness judgments with newness judgments (Kafkas & Montaldi, 2015). Although this study found very few false alarms with oldness judgments, it did, however, find that familiarity hits initially showed larger pupil responses than familiarity misses of newness in the newness judgment task, which might be regarded as equivalent to false alarms in the oldness judgment. In the period leading up to the response, pupil size was the same. Our results suggest that the different procedures are associated with distinct patterns of pupil responses, although this requires further investigation.

We can confidently exclude the possibility that the similar magnitude of pupil dilation between Fs and FAs was the result of different levels of familiarity strength and/or increased cognitive effort applied to FA than F items (or vice versa). We only compared RT-matched trials between Fs and FAs, such that only trials that had similar RTs were included in the analysis. This should have effectively and considerably reduced differences in cognitive effort between contrasting



categories. Although we acknowledge that RT-matching may still be insufficient to perfectly match effort and familiarity strength between any two categories, changes in RTs remain a popular proxy for cognitive effort (Beatty, 1982; Gomes & Mayes, 2015a, 2015b, 2020; Montefinese et al., 2013; Porter et al., 2007; Vö et al., 2008).

Another potential criticism could be raised that the lack of a statistical difference in the pupillary response between Fs and FAs was because either Fs or FAs (or both) simply failed to produce a distinctive pupillary response altogether (i.e., the measures were too noisy). This is, however, inconsistent with both Fs and FAs showing a substantially reduced pupillary response relative to Rs as time elapsed, as well as increased dilation relative to CRs and Ms, both of which are well-documented effects in the literature (Gomes et al., 2015; Kafkas & Montaldi, 2011, 2012; Otero et al., 2011).

Another possibility is that the increase in pupil dilation for FAs reflects a misattribution of "oldness" to the new stimuli. This might occur as the result of participants experiencing an unexpected increase in processing fluency for FAs, which is then interpreted as familiarity (Whittlesea et al., 1990; Whittlesea & Williams, 2000). Old/new decisions are based on whatever cues the participant receives during each trial. For FAs, there is not an objective memory signal as there is for truly old items. Thus, when participants experience increase fluency for FA items, this fluency is unexpected, and gets attributed to prior study.

Indeed, some research has revealed that the neural mechanism responsible for familiarity has a striking resemblance to that of inaccurate memory (Dew & Cabeza, 2013; Gomes et al., 2019). In our experiments, we used very simple line-drawings of objects, so unstudied pictures inevitably shared overlapping features with studied pictures. Furthermore, the relatively shallow encoding conditions (at least during the dot task) resulted in such weak memory traces that, without other cues to help differentiate old from new, likely made processing fluency appealing for guiding recognition memory decisions (Jacoby & Whitehouse, 1989; Whittlesea et al., 1990). Our analysis of image similarity indicated that the similarity between any given new image and studied images was strongly related to increases in pupil size, but only for FAs (Figure 5, left). In addition, the similarity between studied and FA images were overall larger than that between studied and CR images (Figure 5, right), and our logistic regression model was more likely to categorize an image as FA as image similarity increased. In summary, FAs may have led to larger pupils than CRs because FA images shared more perceptual characteristics with study image(s), triggering increased processing fluency during FAs, which was subsequently interpreted as familiarity.

We should stress that we are not arguing that the old/new pupil effect commonly observed in previous studies must reflect this fluency process. In fact, at face value, there appear to be four possibilities: First, accurate and inaccurate

familiarity might lead to similar pupillary responses (both in magnitude and across time) due to the operation of a single-system neural mechanism (Berry et al., 2008, 2012). Second, it is also possible that accurate and inaccurate familiarity involves largely distinct neural mechanisms, but with some overlapping mechanisms that trigger a similar pupillary response. Third, if fluency without familiarity can support above-chance performance in some recognition memory tests (Gomes et al., 2017; Leynes & Zish, 2012; Whittlesea & Leboe, 2003), then the observation of increased pupil dilation in response to Fs, need not be a hallmark of familiarity at all, but could actually indicate contamination from fluency that does not help produce familiarity (Voss, Lucas, et al., 2012). Finally, and perhaps more likely, pupil dilation may indicate familiarity feelings but, when effort and familiarity memory strength are equated, the brain may simply fail to distinguish between accurate and inaccurate familiarity, producing similar pupil dilations. Fluency may represent an intermediate stage following Jacoby's fluency attribution notion, given that our results suggest that inaccurate familiarity relates to how similar false alarm stimuli are to studied stimuli. This indicates that an underlying fluency process contributes to inaccurate familiarity and triggers pupil dilation. Accurate familiarity may also have a contribution from this same fluency process, the extent of which will require further investigation.

Unfortunately, we cannot disentangle these competing possibilities, as our experiments were not tailored to look specifically at the relationship between fluency and familiarity. Nevertheless, it is clear that pupillometry will not be useful in separating these two processes quantitatively with standard recognition memory paradigms. In order to adequately isolate accurate from inaccurate familiarity memory, future pupillometric studies must be able to differentiate F from FA processing, a contrast seldom reported in the relevant literature. One way to achieve this could be, for example, using meaningless stimuli, such as squiggles, which are unlikely to evoke strong fluency effects, but still trigger typical responses associated with familiarity memory (Voss et al., 2012; Voss & Paller, 2007). However, it remains distinctly possible that appropriately matched accurate and inaccurate familiarity are associated with indistinguishable pupillary responses.

## 4.1 | Limitations of the present study

One important limitation of the present study is that measuring similarity using the BOLAR index makes certain assumptions that may be false. When judging similarity, it is clear that different individuals may select different features, use different algorithms, and rank images differently from the BOLAR index. The BOLAR index used to compute similarity in the present study is a purely data-driven method, which computes distances between BOLAR vectors based on the image pixels

(see Methods). It may well be that the resulting difference maps contain features that our brains simply do not use or, just as plausible, may fail to select features that our brain does use. If the overlap is insufficient, then the correlation between objective and subjective similarity may be poor. Averaged subjective similarity ratings between certain stimulus categories have been computed (Frank et al., 2020; Migo et al., 2013) but, unfortunately, subjective estimates of similarity were not acquired in the present study. Future research will need to determine whether the findings reported here with the BOLAR index also occur with subjective similarity ratings.

Another potential concern was the lack of a significant interaction between pupil dilation and BOLAR index when the response category was used as the dependent variable in the logistic multilevel regression analysis. This poses a problem to our account that FAs are driven by perceptual fluency, as we would have expected a significant modulation of fluency (as indexed by the BOLAR values) on pupil dilation for FAs but not CRs. Nevertheless, we did find a modest correlation between the BOLAR index and the pupil dilation for FAs but not CRs at a participant level (see Figure 5, left). Even though the FA effect was admittedly weak, the correlation coefficient for FAs was still significantly different from that of CRs.

Finally, another limitation of the present study is that, for the main pupil results, the data for the three experiments were collapsed despite the fact that two different encoding tasks were used. Collapsing the pupil data had the major advantage of healthily increasing the power of the comparison of F and FA pupil dilation analyses. However, collapsing pupil data across different encoding conditions is questionable if they have produced familiarity memory of different strengths. There is evidence of this as is indicated in the results above. Nevertheless, collapsing would be justified if there were no effect of encoding condition on relative F and FA pupil dilation. No such interaction effects were found with the experiment factor on any of the pupil dilation analyses (see Methods). The limitation arises because comparing the pupil measures of the three experiments would mean that at least some of the statistics were underpowered (particularly for Bayes Factor analysis). Nevertheless, there was no trend towards F and FA pupil differences in the data. In the future, adequately powered analyses should use a wider range of conditions so as to systematically vary familiarity memory strength sufficiently to assess whether F versus FA pupil differences emerge when memory strength becomes greater than we observed.

## 4.2 | Implications of this study to lie detection

Our results suggest that it may be difficult to differentiate accurate from inaccurate (familiarity) memory, as increases

in pupil dilation were observed both when subjects correctly remembered images as well as when they wrongly believed the images were old. Perhaps a germane question could be asked: do the results of the present study have implications in the investigation of deception as measured with pupillometry?

Several studies have reported increases in pupil diameter when participants were asked to lie (Dionisio et al., 2001; Heilveil, 1976; Trifiletti et al., 2020; Wang et al., 2010), which has called for the development of lie detection technology based on pupillometry. Even though these studies showed greater increases in pupil dilation for deceptive than truthful answers, it is unclear if the distinction between deceptive and inaccurate memory during recognition memory tests can be as easily detected with pupillometry. It will, therefore, be important to ascertain if actively lying about memories of particular items or events produces greater pupil dilation than inaccurate memory to those same items or events during recognition. If the pupillary response to lying is found to be no different than when participants wrongly believe they remembered an item, this may limit the utility of pupillometry as a lie detector. Future research may help find a more conclusive answer.

## AUTHOR CONTRIBUTIONS

**Carlos Alexandre Gomes:** Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing-original draft; Writing-review & editing. **Daniela Montaldi:** Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing-review & editing. **Andrew Mayes:** Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing-review & editing.

## ORCID

Carlos A. Gomes  <https://orcid.org/0000-0001-5604-4377>

## REFERENCES

- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276–292. <https://doi.org/10.1037/0033-2909.91.2.276>
- Berry, C. J., Shanks, D. R., & Henson, R. N. A. (2008). A Single-system account of the relationship between priming, recognition, and fluency. *Journal of Experimental Psychology: Learning Memory and Cognition*, 34(1), 97–111. <https://doi.org/10.1037/0278-7393.34.1.97>
- Berry, C. J., Shanks, D. R., Speekenbrink, M., & Henson, R. N. A. (2012). Models of recognition, repetition priming, and fluency: Exploring a new framework. *Psychological Review*, 119(1), 40–79. <https://doi.org/10.1037/a0025464>

- Brocher, A., & Graf, T. (2017). Decision-related factors in pupil old/new effects: Attention, response execution, and false memory. *Neuropsychologia*, 102, 124–134. <https://doi.org/10.1016/j.neuropsychologia.2017.06.011>
- Dew, I. T. Z., & Cabeza, R. (2013). A broader view of perirhinal function: From recognition memory to fluency-based decisions. *Journal of Neuroscience*, 33(36), 14466–14474. <https://doi.org/10.1038/jes.2010.44>
- Dionisio, D. P., Granholm, E., Hillix, W. A., & Perrine, W. F. (2001). Differentiation of deception using pupillary responses as an index of cognitive processing. *Psychophysiology*, 38(2), 205–211. <https://doi.org/10.1017/S0048577201990717>
- Frank, D., Gray, O., & Montaldi, D. (2020). SOLID-Similar object and lure image database. *Behavior Research Methods*, 52(1), 151–161. <https://doi.org/10.3758/s13428-019-01211-7>
- Golding, S. D., & Pape, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, 21, 90–95. <https://doi.org/10.1177/0963721412436811>
- Gomes, C. A., Figueiredo, P., & Mayes, A. (2016). Priming for novel object associations: Neural differences from object item priming and equivalent forms of recognition. *Hippocampus*, 26(4), 472–491. <https://doi.org/10.1002/hipo.22537>
- Gomes, C. A., & Mayes, A. (2015a). Does long-term object priming depend on the explicit detection of object identity at encoding? *Frontiers in Psychology*, 6, 270. <https://doi.org/10.3389/fpsyg.2015.00270>
- Gomes, C. A., & Mayes, A. (2015b). The kinds of information that support novel associative object priming and how these differ from those that support item priming. *Memory*, 23(6), 901–927. <https://doi.org/10.1080/09658211.2014.937722>
- Gomes, C. A., & Mayes, A. (2020). Study–test congruence of response levels in item stimulus–response priming. *Memory and Cognition*, 48(5), 839–855. <https://doi.org/10.3758/s13421-020-01021-9>
- Gomes, C. A., Mecklinger, A., & Zimmer, H. (2017). Behavioural and neural evidence for the impact of fluency context on conscious memory. *Cortex*, 92, 271–288. <https://doi.org/10.1016/j.cortex.2017.04.008>
- Gomes, C. A., Mecklinger, A., & Zimmer, H. (2019). The neural mechanism of fluency-based memory illusions: The role of fluency context. *Learning & Memory*, 26(2), 61–65. <https://doi.org/10.1101/lm.048637.118>
- Gomes, C. A., Montaldi, D., & Mayes, A. (2015). The pupil as an indicator of unconscious memory: Introducing the pupil priming effect. *Psychophysiology*, 52(6), 754–769. <https://doi.org/10.1111/psyp.12412>
- Heaver, B., & Hutton, S. B. (2011). Keeping an eye on the truth? Pupil size changes associated with recognition memory. *Memory*, 19(4), 398–405. <https://doi.org/10.1080/09658211.2011.575788>
- Heilveil, I. (1976). Deception and pupil size. *Journal of Clinical Psychology*, 32(3), 675–676. [https://doi.org/10.1002/1097-4679\(197607\)32:3<675::AID-JCLP2270320340>3.0.CO;2-A](https://doi.org/10.1002/1097-4679(197607)32:3<675::AID-JCLP2270320340>3.0.CO;2-A)
- Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, 118(2), 126–135. <https://doi.org/10.1037/0096-3445.118.2.126>
- Johnston, W. A., Dark, V. J., & Jacoby, L. L. (1985). Perceptual fluency and recognition judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(1), 3–11. <https://doi.org/10.1037/0278-7393.11.1.3>
- Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology*, 64(10), 1971–1989. <https://doi.org/10.1080/17470218.2011.588335>
- Kafkas, A., & Montaldi, D. (2012). Familiarity and recollection produce distinct eye movement, pupil and medial temporal lobe responses when memory strength is matched. *Neuropsychologia*, 50(13), 3080–3093. <https://doi.org/10.1016/j.neuropsychologia.2012.08.001>
- Kafkas, A., & Montaldi, D. (2015). The pupillary response discriminates between subjective and objective familiarity and novelty. *Psychophysiology*, 52(10), 1305–1316. <https://doi.org/10.1111/psyp.12471>
- Kuchinke, L., Trapp, S., Jacobs, A. M., & Leder, H. (2009). Pupillary responses in art appreciation: Effects of aesthetic emotions. *Psychology of Aesthetics, Creativity, and the Arts*, 3(3), 156–163. <https://doi.org/10.1037/a0014464>
- Laeng, B., Waterloo, K., Johnsen, S. H., Bakke, S. J., Låg, T., Simonsen, S. S., & Høgsæt, J. (2007). The eyes remember it: Oculography and pupillometry during recollection in three amnesic patients. *Journal of Cognitive Neuroscience*, 19(11), 1888–1904. <https://doi.org/10.1162/jocn.2007.19.11.1888>
- Leynes, A. P., & Zish, K. (2012). Event-related potential (ERP) evidence for fluency-based recognition memory. *Neuropsychologia*, 50(14), 3240–3249. <https://doi.org/10.1016/j.neuropsychologia.2012.10.004>
- Mayes, A., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, 11(3), 126–135. <https://doi.org/10.1016/j.tics.2006.12.003>
- Migo, E. M., Montaldi, D., & Mayes, A. R. (2013). A visual object stimulus database with standardized similarity information. *Behavior Research Methods*, 45(2), 344–354. <https://doi.org/10.3758/s13428-012-0255-4>
- Mill, R. D., O'Connor, A. R., & Dobbins, I. G. (2016). Pupil dilation during recognition memory: Isolating unexpected recognition from judgment uncertainty. *Cognition*, 154, 81–94. <https://doi.org/10.1016/j.cognition.2016.05.018>
- Montaldi, D., Spencer, T. J., Roberts, N., & Mayes, A. R. (2006). The neural system that mediates familiarity memory. *Hippocampus*, 16(5), 504–520. <https://doi.org/10.1002/hipo.20178>
- Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). The “subjective” pupil old/new effect: Is the truth plain to see? *International Journal of Psychophysiology*, 89(1), 48–56. <https://doi.org/10.1016/j.ijpsycho.2013.05.001>
- Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, 48(10), 1346–1353. <https://doi.org/10.1111/j.1469-8986.2011.01217.x>
- Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *Quarterly Journal of Experimental Psychology*, 60(2), 211–229. <https://doi.org/10.1080/17470210600673818>
- Spencer, T. J., Montaldi, D., Gong, Q. Y., Roberts, N., & Mayes, A. R. (2009). Object priming and recognition memory: Dissociable effects in left frontal cortex at encoding. *Neuropsychologia*, 47(13), 2942–2947. <https://doi.org/10.1016/j.neuropsychologia.2009.06.022>
- Trifiletti, E., D'Ascenzo, S., Lugli, L., Cocco, V. M., Di Bernardo, G. A., Iani, C., Rubichi, S., Nicoletti, R., & Vezzali, L. (2020). Truth and lies in your eyes: Pupil dilation of White participants in truthful and deceptive responses to White and Black partners.



- PLoS ONE, 15(10), e0239512. <https://doi.org/10.1371/journal.pone.0239512>
- Vö, M. L. H., Jacobs, A. M., Kuchinke, L., Hofmann, M., Conrad, M., Schacht, A., & Hutzler, F. (2008). The coupling of emotion and cognition in the eye: Introducing the pupil old/new effect. *Psychophysiology*, 45(1), 130–140. <https://doi.org/10.1111/j.1469-8986.2007.00606.x>
- Voss, J. L., Federmeier, K. D., & Paller, K. A. (2012). The potato chip really does look like Elvis! Neural hallmarks of conceptual processing associated with finding novel shapes subjectively meaningful. *Cerebral Cortex*, 22(10), 2354–2364. <https://doi.org/10.1093/cercor/bhr315>
- Voss, J. L., Lucas, H. D., & Paller, K. A. (2012). More than a feeling: Pervasive influences of memory without awareness of retrieval. *Cognitive Neuroscience*, 3(3–4), 193–207. <https://doi.org/10.1080/17588928.2012.674935>
- Voss, J. L., & Paller, K. A. (2007). Neural correlates of conceptual implicit memory and their contamination of putative neural correlates of explicit memory. *Learning and Memory*, 14(4), 259–267. <https://doi.org/10.1101/lm.529807>
- Wang, J. T. Y., Spezio, M., & Camerer, C. F. (2010). Pinocchio's pupil: Using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *American Economic Review*, 100(3), 984–1007. <https://doi.org/10.1257/aer.100.3.984>
- Wang, W. C., Brashier, N. M., Wing, E. A., Marsh, E. J., & Cabeza, R. (2016). On known unknowns: Fluency and the neural mechanisms of illusory truth. *Journal of Cognitive Neuroscience*, 28(5), 739–746. [https://doi.org/10.1162/jocn\\_a\\_00923](https://doi.org/10.1162/jocn_a_00923)
- Whittlesea, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19(6), 1235–1253. <https://doi.org/10.1037/0278-7393.19.6.1235>
- 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(6), 716–732. [https://doi.org/10.1016/0749-596X\(90\)90045-2](https://doi.org/10.1016/0749-596X(90)90045-2)
- Whittlesea, B. W. A., & Leboe, J. P. (2003). Two fluency heuristics (and how to tell them apart). *Journal of Memory and Language*, 49(1), 62–79. [https://doi.org/10.1016/S0749-596X\(03\)00009-3](https://doi.org/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(3), 547–565. <https://doi.org/10.1037/0278-7393.26.3.547>
- Zelinsky, G. J. (2003). Detecting changes between real-world objects using spatiochromatic filters. *Psychonomic Bulletin and Review*, 10(3), 533–555. <https://doi.org/10.3758/BF03196516>

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the Supporting Information section.

ANOVA results with non-matched data

**How to cite this article:** Gomes CA, Montaldi D, Mayes A. Can pupillometry distinguish accurate from inaccurate familiarity?. *Psychophysiology*.

2021;58:e13825. <https://doi.org/10.1111/psyp.13825>