Open Research Online

The Open University's repository of research publications and other research outputs

You can believe your eyes: measuring implicit recognition in a lineup with pupillometry

Journal Item

How to cite:

Elphick, Camilla E. J.; Pike, Graham E. and Hole, Graham J. (2020). You can believe your eyes: measuring implicit recognition in a lineup with pupillometry. *Psychology, Crime and Law*, 26(1) pp. 67–92.

For guidance on citations see [FAQs](#).

© 2019 Informa UK Limited, trading as Taylor Francis Group

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Accepted Manuscript

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.1080/1068316X.2019.1634196>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

You can Believe your Eyes: Measuring Implicit Recognition in a Lineup with Pupillometry.

Camilla E. J. Elphick^a, Graham E. Pike^b, Graham J. Hole^a

^aUniversity of Sussex, Falmer, Brighton BN1 9RH

C.Elphick@sussex.ac.uk (corresponding author)

G.Hole@sussex.ac.uk

^bOpen University, Walton Hall, Kents Hill, Milton Keynes MK7 6AA

graham.pike@open.ac.uk

There is no conflict of interest.

ACKNOWLEDGEMENTS

We thank Sam Hutton and SR Research for their excellent hardware and software support

Abstract

As pupil size is affected by cognitive processes, we investigated whether it could serve as an independent indicator of target recognition in lineups. Participants saw a simulated crime video, followed by two viewings of either a target-present or target-absent video lineup while pupil size was measured with an eye-tracker. Participants who made correct identifications showed significantly larger pupil sizes when viewing the target compared with distractors. Some participants were uncertain about their choice of face from the lineup, but nevertheless showed pupillary changes when viewing the target, suggesting covert recognition of the target face had occurred. The results suggest that pupillometry might be a useful aid in assessing the accuracy of an eyewitness' identification.

Keywords: pupillometry, eyewitness identification, covert recognition, face processing, credibility

Measuring Recognition in Lineups with Pupillometry

In contrast to our skill in recognising familiar faces, unfamiliar face recognition is difficult and error-prone (review in Hancock, Bruce, & Burton, 2000). For example, in a study using CCTV images and comparison photographs, Bruce, Henderson, Newman and Burton (2001) reported that matching accuracy was about 75% for unfamiliar faces, compared to approximately 90% for familiar faces. Eyewitness misidentifications may be a major factor in miscarriages of justice (the Innocence Project, 2017; see also Dwyer, Neufeld, & Scheck, 2000; Wells & Olson, 2003).

Given the unreliability of eyewitnesses' conscious responses to lineups, research has investigated alternative methods for assessing their credibility (e.g. MacLin, MacLin, & Malpass, 2001; Wright & Stroud, 2002). For instance, eye-tracking has been used to distinguish between participants who made an identification from those who did not in simultaneous lineups (where all the faces are seen at the same time), but this technique is not helpful for differentiating correct identifications from false identifications in sequential lineups (when pictures or video clips are shown one at a time: Mansour & Flowe, 2010).

Using ERPs (event-related potentials), Lefebvre, Marchand, Smith and Connolly (2007) found that participants who made correct identifications showed an increased P300 response to the target compared to distractors. The P300 was also significantly larger for correct identifications than for misidentifications. Unfortunately, it would currently be impractical to measure ERPs in a real-world setting.

Attempting to compare scores on established face recognition tests with lineup responses, Bindemann, Brown, Koyas and Russ (2012) found that a face recognition test postdicted lineup performance in participants who made an identification, but not in

Measuring Recognition in Lineups with Pupillometry

participants who misidentified the target, missed the target, or correctly rejected all faces in a target-absent condition. In a follow-up study, the face recognition test only postdicted lineup performance for correct rejections. Both experiments (across target-present and target-absent lineups) indicated that the face recognition test provided a good index of eyewitness reliability for participants who made an identification, but not for those who made no identification.

Confidence has also been studied with regards to whether it can be diagnostic of identification accuracy, and judges and juries attach considerable weight to a witness' confidence when evaluating them (e.g. Wells, Ferguson, & Lindsay, 1981). Although not all research has found confidence ratings to be a reliable guide to identification accuracy (Brewer & Burke, 2002; Cutler, Penrod, & Stuve, 1988), more recent research suggests that the confidence-accuracy relationship is stronger than previously thought (Wixted, Read, & Lindsay, 2016). If confidence can be recorded immediately after an identification is made, including before any feedback is provided, then it may be a useful indicator in instances where the witness makes a selection from the lineup (Sauer, Brewer, Zweck, & Weber, 2009; Sauerland & Sporer, 2009). Research has also shown that very confident eyewitnesses tend to have relatively high degrees of accuracy, whereas the same is not true of unconfident witnesses (Brewer & Palmer, 2010). However, one problem with confidence ratings is that they are susceptible to being influenced by post-identification feedback, especially if made retrospectively rather than immediately after the lineup has taken place (Wells & Bradfield, 1998; Wells et al., 1998).

Another method of achieving the same goal is to use pupillometry. Pupil size is not determined solely by ambient luminance, but can be influenced by mental processing, such as

Measuring Recognition in Lineups with Pupillometry

cognitive load: the greater the mental workload, the larger the pupil size (Beatty, 1982; Jainta & Baccino, 2010; Zekveld, Heslenfeld, Johnsrude, Versfeld, & Kramer, 2014; & see Ayres & Paas, 2012; Goldinger & Papesh, 2012; Murphy, Groeger, & Greene, 2016, for reviews).

Pupil size has also been associated with affective processing, as pupils are larger when presented with emotional stimuli than with neutral stimuli (e.g. Bradley, Miccoli, Escrig, & Lang, 2008; Partala & Surakka, 2003; Prehn, Heekeren, & van der Meer, 2011; Snowden et al., 2016; Vö et al., 2008). Pupillometry has also proved useful in indexing memory strength, as pupils have been shown to be larger when retrieving items associated with greater memory strength (Brocher & Graf, 2016; Goldinger & Papesh, 2012; Otero, Weekes, & Hutton, 2011; Papesh, Goldinger, & Hout, 2012), and they also appear to reflect the experience of recognition (Otero et al., 2011). Therefore, pupil size may also reflect the strength of recognition evidence (Montefinese, Vinson, & Ambrosini, 2018). Pupillometry also appears to be useful for measuring implicit memory, as pupillary changes occur in the absence of an overt response (van Rijn, Dalenberg, Borst, & Sprenger, 2012), and can even occur despite efforts to deceive (Heaver & Hutton, 2011).

Pupillometry has seldom been used in face recognition research. However, Goldinger, He, and Papesh (2009) have shown that pupil sizes were larger when looking at other-race faces than own-race faces. Considering the social importance of faces, and combining this with the findings that pupils respond to memory strength, it appears that pupillary changes could be a reliable measure of face recognition, and applicable to eyewitness lineup procedures (review in Goldinger & Papesh, 2012). However, as far as we know, pupil size has not been investigated within the context of a lineup until now.

Measuring Recognition in Lineups with Pupillometry

Pupillometry has the potential to be a useful supplementary measure of eyewitness identification performance. It would be desirable to have a measure of eyewitness performance that is independent of the explicit decision processes involved in making an identification. Previous attempts to assess eyewitness accuracy, using measures of witness confidence or generalised face recognition ability, do not fulfil this criterion as these are alternative explicit measures of recognition that may well be contaminated by the conscious decision processes involved in making an identification in the first place. In contrast, pupil size is a physiological response outside of the witness' conscious control, and hence more likely to be independent of their overt decision about the lineup. Unlike traditional measures, pupillometry also reveals real-time fluctuations in cognitive processing, such as changes in mental effort, engagement and memory strength *while viewing a suspect's face*. Pupillary measures may thus provide more nuanced information about eyewitness identification performance, which could be used to assist decisions about an eyewitness' credibility.

The present study investigated whether pupil size is a good predictor of lineup identification accuracy, using both target-present and target-absent lineups. Target-present lineups were employed to determine whether pupil size changes can be used to assess whether a participant is viewing a face they have seen previously (i.e. whether pupil size reliably discriminates between a target face and distractors). Target-absent lineups were also employed, as it is important to know what happens to pupil sizes if the perpetrator of the crime is *not* present in the lineup. To be an effective method, pupillometry would need to consistently differentiate target faces from distractors in target-present lineups *and* show that pupil size changes are not associated consistently with any single face in the lineup, especially when the target is absent.

Experiment 1

Methods

Participants

In the target-present condition, 51 participants with normal or corrected-to-normal vision were recruited in exchange for university course credits or cash. Two participants were subsequently excluded for making multiple identifications in both lineups, contrary to instructions. This left 49 participants aged between 18 and 26 ($M = 19.61$, $SD = 1.72$) for the analysis (15 males and 34 females). Participants were recruited until there were at least ten for each category of identification response (Identifiers, Non-identifiers and Misidentifiers) in each lineup¹. In the target-absent condition there were 26 participants (2 males, 24 females), aged between 18 and 35 ($M = 20.42$, $SD = 3.23$).

Apparatus and Stimuli

First, participants saw a silent video clip of a staged non-violent crime, in which a man attempted to steal another man's bag. After a brief altercation, the two men ran off, fighting over the bag. The video lasted one minute. It was recorded in .wmv format (768 x

¹ Ordinarily, the number of participants per condition is under the control of the experimenter. However, in this case, allocation to a condition depended entirely on the participants' identification decision. Our strategy was therefore adopted merely to ensure that we had a reasonable sample size for each type of decision, even those that occurred relatively infrequently. A pilot study showed that we were able to obtain reasonably reliable effects with only eight participants per condition, so our choice to keep testing until we had at least ten participants for each type of identification response was motivated by this consideration.

Measuring Recognition in Lineups with Pupillometry

576 pixels) and converted to XVID for compatibility with the eye-tracking software, Experiment Builder (SR Research, n.d.).

Two lineup conditions were used: target-present and target-absent. In the target-present condition, the lineup presented ten colour video clips of head and shoulders against a white background (nine distractors and one target). In each clip, the individual faced the camera before turning their head slowly to the right, to the left, and back to centre. These were constructed by the VIPER Unit of the West Yorkshire Police, using the VIPER (n.d.) video identification parade database. Trained officers selected distractors from the VIPER database (containing thousands of faces) to match a basic verbal description of the target (ensuring that distractors were also a reasonable visual match to the physical appearance of the target). Videos were cropped and matched for size (17.5 x 13.3 cm), resolution (768 x 576 pixels), time (12 seconds), and luminance using an app (myLightMeterPro, 2015). The blinds were drawn to control lighting levels. In the target-absent condition, the target face was removed from the lineup, so participants only saw the 9 distractor faces, so that results were not confounded by seeing an additional novel face².

Experiment Builder was run using a 21.5 inch iMac computer, and a desktop Eyelink 1000 eye-tracker (SR Research, n.d.) that recorded pupil position and size using infrared

² We were faced with two options, either to introduce an additional face to replace the target, or to delete the target. We thought that deleting the target was the lesser of two evils, as it meant that the only difference between the TP and TA lineups was that the TA lineup had one less face, and everything else was kept constant. Introducing a novel face runs the risk of changing the nature of the lineup. In experiment three, for practical reasons (increased speed of testing in a real-life situation) we used a different lineup size from the previous two experiments (and a different target person) and obtained similar effects, suggesting that lineup size (within the limits of those we used) was not a major influence on our results, although future studies should attempt to keep lineup size constant.

Measuring Recognition in Lineups with Pupillometry

illumination. The participant's head was stabilised with a chin rest, at approximately 60 cm from the computer screen.

Design

This study used a mixed design: independent measures on *identification response* (with three levels: Identifiers, Non-identifiers, and Misidentifiers) and repeated measures on *face type* (with three levels: Pre-target distractors, Target, and Post-target distractors). With target-present lineups, the target was the person seen in the staged crime video.

[Insert Fig. 1 here]

With target-absent lineups, the "target" was the misidentified distractor who had not been encountered before the lineup (known as a "false positive"). The dependent variable was pupil size, calculated as a percentage of each participant's overall pupil size range during the experiment. The procedure is described in detail in the results.

Procedure

Participants were briefed before placing their chins in the chin rest, and their eye movements were calibrated to nine points. After reading instructions, their gaze was monitored with a drift check. Drift checks monitor the accuracy of eye-tracking data and involve looking at a black dot on a white screen. Then the simulated crime video was played. This was followed by a filler task, the long version of the Glasgow Face Matching Task (GFMT) (Burton, White, & McNeill, 2010). This was chosen to reflect the fact that eyewitnesses see many faces between a real crime and the lineup, so we considered it to be

Measuring Recognition in Lineups with Pupillometry

more ecologically valid than isolating participants from any faces between the crime video and the lineup in the experiment.

A pilot version of this experiment revealed that participants who misidentified a distractor tended to do so when looking at the first face in the lineup, suggesting that they would benefit from becoming acquainted with the task before viewing the lineup itself. Therefore, in the main experiment, immediately after the filler task and before seeing the lineup, participants were given practice at identifying a face: they were shown two faces that did not resemble the target. These were filmed in the same way as the other lineup faces, and participants were told explicitly that they were for practice only.

After completing the practice session, participants saw a video lineup relating to the staged crime. Lineup video clips were displayed one at a time with a drift check between adjacent clips. For each clip, the participant was asked to click 'Y' (yes) if they thought the face was the target and 'N' (no) if they thought it was a distractor. They were asked not to press 'Y' more than once per lineup. Each clip played for 12 seconds regardless of how quickly the participant responded, but the ISI was not fixed, as each clip was separated by a drift check. This took approximately 2-3 seconds: as soon as their eye stabilised, the next trial began. Each participant saw the lineup video clips in a different pseudo-random order (in target-present lineups, the target was never first or last in the lineup). After the final clip, the lineup procedure was repeated, with the clips displayed in a different pseudo-random sequence, and participants were asked to make another Y/N response. They were told that this response could be the same as it had been in the first lineup, or that they could make a different response if they wished. The eye-tracker recorded pupillary responses and identification decisions as participants viewed the clips.

Results

Target-Present condition.

To standardise pupil size measurements between participants, the following procedure was used. For each participant, the eye-tracker produced a mean pupil size for each video clip in the lineup. We subtracted each participant's smallest mean from their largest mean, to produce a difference score. The mean pupil size for each clip was then expressed as a percentage of this difference.

From these values, three pupil size measures were produced for each participant. First, pupil sizes for distractors seen *before* the target were averaged together to produce a single mean pupil size measure, labelled “pre-target distractors”. There was only one target, so only one measure was available for each participant, labelled “target”. Finally, pupil sizes for distractors seen *after* the target were averaged together to give a single mean pupil size measure labelled “post-target distractors”.

To determine whether pupil sizes changed in response to the target, two two-way mixed ANOVAs were conducted, one for the first lineup and another for the second. For each ANOVA, there was one within-subjects factor, *face type* (with three levels: pre-target distractors, target, and post-target distractors) and one between-subjects factor: *identification response* (with three levels: identifiers, non-identifiers, and misidentifiers). "Identifiers" were participants who correctly identified the target; "non-identifiers" mistakenly thought the target was absent; and "misidentifiers" mistook a distractor for the target.

Measuring Recognition in Lineups with Pupillometry

Some participants made multiple misidentifications in one of the lineups but not the other. Their data were only analysed for the lineup in which they performed as instructed. This left forty-five participants for the first lineup (22 identifiers, 13 non-identifiers, and 10 misidentifiers).

Fig. 2 suggests that, for the first lineup, pupil sizes were larger for the target compared to distractors, and different between the three identification response groups. This interpretation was supported by the ANOVA results for the first lineup, which showed significant effects of *face type*, $F(2,84) = 26.22, p < .001, r = .48, \eta^2 = .38$ (Pre: $M = 59.80, SE = 2.50$; Target: $M = 75.40, SE = 2.90$; Post: $M = 50.80, SE = 2.70$), and *identification response*, $F(2,42) = 4.02, p = .03, r = .29, \eta^2 = .16$ (Identifiers: $M = 66.90, SE = 2.50$; Misidentifiers: $M = 54.50, SE = 3.70$; Non-identifiers: $M = 64.60, SE = 3.20$), and a significant interaction between *face type* and *identification response*, $F(4,84) = 2.49, p < .05$.

[Insert Fig. 2 here]

Planned contrasts revealed that pupil sizes for *identifiers* were significantly larger for the target than for pre-target distractors, $F(1,21) = 96.62, p < .001, r = .89, \eta^2 = .82$ (29% difference in size) and post-target distractors $F(1,21) = 96.33, p < .001, r = .91, \eta^2 = .82$ (33% difference in size). For *non-identifiers*, pupils were significantly larger for the target than for post-target distractors, $F(1,12) = 9.99, p = .01, r = .67, \eta^2 = .45$ (21%), but not for pre-target distractors $F(1,12) = 1.79, p = .21$ (9%). There were no significant comparisons for *misidentifiers*: pre-target distractors, $F(1,9) = 1.05, p = .33$ (8%) or post-target distractors $F(1,9) = 4.32, p = .07$ (20%).

Measuring Recognition in Lineups with Pupillometry

Between identifiers, non-identifiers and misidentifiers, three one-way ANOVAs revealed significant pupillary differences for the target, $F(2,44) = 6.21, p = .01, r = .33$, but not for pre-target distractors, $F(2,44) = 1.16, p = .33$, or post-target distractors, $F(2,44) = 1.52, p = .23$. Planned comparisons revealed that when looking at the target, there were significantly different differences between *identifiers* and *non-identifiers* ($p < .05$) and between *identifiers* and *misidentifiers* ($p = .01$), but not between *misidentifiers* and *non-identifiers* ($p = .16$).

We then conducted a paired samples *t*-test for each identification response, to compare pupil sizes when people looked at the target compared to when looking at the distractor that elicited the largest pupil size (LD). As presented in Table 1., for *identifiers* alone, there was no significant difference in pupil size when looking at the two face types, $t(31) = 0.76, p = .46$. However, in *misidentifiers* and *non-identifiers*, pupil sizes were significantly larger when looking at the LD compared to when looking at the target: *misidentifiers*, $t(9) = 3.02, p = .0$; *non-identifiers*, $t(12) = 2.80, p = .02$.

[Insert Table 1 here]

For the second lineup, data from 44 participants were analysed. As shown in fig. 2, there was a significant effect of *face type*, $F(2,82) = 12.28, p < .001, r = .36, \eta^2 = .23$ (Pre: $M = 41.70, SE = 2.70$; Target: $M = 51.40, SE = 3.90$; Post: $M = 32.70, SE = 2.90$), but no effect of *identification response*, $F(2,41) = 0.25, p = .77$, (Identifiers: $M = 42.60, SE = 3.20$; Misidentifiers: $M = 39.40, SE = 4.70$; Non-identifiers: $M = 43.80, SE = 4.30$). There was also no interaction between *face type* and *identification response*, $F(4,82) = 0.92, p = .46$. As

Measuring Recognition in Lineups with Pupillometry

there was no significant effect of *identification response* and no interaction. we did not conduct *t*-tests as we had done for the first lineup.

Thus, pupillary responses discriminated between face types and between identification decisions first time participants saw the lineup, but not when they viewed it a second time.

Using pupil size to predict identification response:

Two binary logistic regressions were used to determine whether pupil size change could predict whether participants made a correct or incorrect lineup decision. The predictor variable was pupil size (the mean difference between the target and the distractors) and the outcome variable was identification accuracy (correct or incorrect).

For the first lineup, the logistic regression was statistically significant, $\chi^2(1) = 6.49$ $p = .01$. The model explained 17.6% (Nagelkerke R^2) of the variance in lineup decision outcome (i.e. whether or not participants were correct) and correctly classified 69.6% of cases. For the first lineup, pupil size was therefore a fairly good measure of identification performance. This was not true for the second lineup, for which the logistic regression was not significant, $\chi^2(1) = 0.50$, $p = .48$.

Target-Absent condition.

Participants were divided into two categories based on their identification response: "misidentifiers", participants who mistook a distractor for the target, and "correct rejectors", who correctly responded that the target was not present. We tested until there was a minimum

Measuring Recognition in Lineups with Pupillometry

of ten participants in each identification response group, which resulted in 26 participants (10 misidentifiers and 16 correct rejectors).

Three pupil size measures were taken from each participant. In the absence of a target, for *misidentifiers* we calculated pupillary responses to the misidentified face and treated this as the ‘target’, but called it the “false positive”. Pupil sizes for distractors seen before it were averaged together to produce a single measure, labelled “pre-false positive faces (Pre)”. There was only one false positive, so only one measure was available for each participant, labelled “false positive”. Finally, pupil sizes for distractors seen after the false positive were averaged together to give a single measure labelled “post-false positive faces (Post)”. For participants who *correctly rejected* all faces, we did not have a false positive face to compare to the target. Therefore, we selected the distractor that had been misidentified most often by the misidentifiers (47% of the time) and designated it to be the “false positive”, as we considered it most likely to be considered a close match to the target and therefore most likely to elicit a large pupil size. Then, we followed the same procedure that we had followed for the target-absent misidentifiers.

To test whether pupil sizes changed in response to a misidentified face in the absence of the target, two two-way mixed ANOVAs were used for analysis of pupil size measures in response to each lineup. For each, there was one within-subjects factor, *face type* (with three levels: pre, false positive, and post) and one between-subjects factor: *identification response* (with two levels: correct rejectors and misidentifiers).

As shown in fig. 3, for the first lineup there was a significant effect of *face type*, $F(2,48) = 5.92$, $p = .01$, $r = .44$, $\eta^2 = .20$ (Pre: $M = 61.00$, $SE = 3.80$, False positive: $M = 69.40$,

Measuring Recognition in Lineups with Pupillometry

$SE = 4.60$, Post: $M = 50.20$, $SE = 3.90$), but there was no effect of *identification response*, $F(1,24) = 0.80$, $p = .38$ (Misidentifiers: $M = 62.30$, $SE = 4.00$; Correct rejectors: $M = 57.70$, $SE = 3.20$). There was also no interaction between *face type* and *identification response*, $F(2, 48) = 0.36$, $p = .70$.

[Insert fig. 3 here]

Again, we conducted two paired samples t -tests for each identification group, to compare pupil sizes when people looked at the false positive and at the face that elicited the largest pupil size (LF). For *misidentifiers*, there was no significant difference in pupil size between the face types, $t(9) = 1.89$, $p = .09$; but for *Correct rejectors*, there was a significant difference, $t(15) = 5.52$, $p = .01$.

[Insert Table 2 here]

Inspection of fig. 3, for the second lineup, revealed no significant effects of *face type*, $F(2,48) = 2.61$, $p = .08$ (Pre: $M = 39.10$, $SE = 3.50$; False positive: $M = 43.90$, $SE = 4.80$; Post: $M = 32.80$, $SE = 3.30$) or *identification response*, $F(1,24) = 0.50$, $p = .49$ (Misidentifiers: $M = 40.50$, $SE = 3.40$; Correct rejectors: $M = 36.70$, $SE = 4.20$). Finally, there was no interaction between *face type* and *identification response*, $F(2, 48) = 0.23$, $p = .80$.

As there were no significant effects, we did not conduct t -tests.

Using pupil size to predict identification response:

Two binary logistic regressions were used to determine whether pupil size change could predict whether participants made a correct or incorrect lineup decision. The predictor

Measuring Recognition in Lineups with Pupillometry

variable was pupil size (the mean difference between the false positive and the other faces) and the outcome variable was identification accuracy (correct or incorrect).

The logistic regressions were not statistically significant for either lineup: first lineup, $\chi^2(1) = 1.20$ $p = .27$; second lineup, $\chi^2(1) = 0.23$, $p = .88$. Pupil size did not predict whether participants would correctly reject or misidentify a face.

Discussion

Our principal finding is that pupil size changed in response to the target in target-present lineups. This only occurred in participants who correctly identified the target, and only the first time that they saw him. These pupillary responses also *predicted* identification of the target the first time participants saw him. In contrast, in the target-absent condition pupillary responses did not differentiate between those who correctly rejected the faces and those who misidentified a face. This suggested that it was only the presence of a previously-seen face that resulted in the pupillary effects in the target-present condition.

This novel approach to measuring implicit recognition in a lineup with pupillometry is in line with previous research suggesting that pupillary changes are associated with memory strength (Brocher & Graf, 2016; Otero et al., 2011; Papesh et al., 2012). Our results indicate that when people recognised the target in a TP lineup, their pupils became larger in response to his face. However, this was only found for the first TP lineup. In the second TP lineup all the faces had been seen before, so it may be that different cognitive requirements, such as source monitoring (being able to place the correct face to the crime) (Conway & Dewhurst, 1995) had also affected pupil sizes.

Measuring Recognition in Lineups with Pupillometry

Pupil size was also able to predict identification of the target. In the first TP lineup, pupil size change explained just under 18% of the variance in lineup decision outcome and correctly classified just under 70% of cases. As expected, in the second TP lineup pupil size was not a good predictor of identification accuracy. The same analyses for the target-absent condition found that pupils did not predict the explicit identification responses. Thus, we concluded that the pupillary changes that had occurred in the target-present condition had done so specifically in response to the target, and probably reflected memory strength for the target's face.

Interestingly, the pupillary changes for the second lineup in the target-present condition appeared to be as indistinct as those for the target-absent condition. This suggests that the pupillary changes that occur when viewing the target only occur when the distractors are novel (in the first TP lineup). Once all the faces have been seen, the familiarity distinction between the target and the distractors disappears, meaning that the pupillary changes also disappear. It seems to be the distinction between entirely-novel faces and a previously-seen face that elicits pupillary changes in participants who remember that face.

We wanted to assess pupillometry in relation to traditional measures of lineup identification that rely on explicit decisions. Identification decisions require eyewitnesses to choose between identifying a face or not. This measure is supposed to be based upon their recognition of the suspect as the perpetrator. To do this, the eyewitness first needs to weigh up the options against their memory, and then make an explicit choice. However, these decisions rely on conscious processes that might mask implicit recognition in some cases, for instance if implicit recognition has not reached the threshold required for an identification, if an eyewitness does not want to make an error (for fear of wrongful conviction), or if they

Measuring Recognition in Lineups with Pupillometry

actually know the perpetrator and make a decision not to identify them. Pupillometry appears to overcome these shortcomings, as pupillary changes occur in the absence of an overt response (van Rijn et al., 2012), and can occur despite efforts to deceive (Heaver & Hutton, 2011). Thus, pupil size appears to reflect an implicit response such as memory strength that is independent of explicit responses.

UK police lineup procedures use a sequential lineup of nine video clips, with the sequence presented twice (Seale-Carlisle & Mikes, 2016). Each face in the lineup is numbered and eyewitnesses make their identification decision at the end of both lineups. Our second experiment explored pupillary responses using this paradigm with three additional measures. We asked participants to make an identification response once for each lineup, so that we could compare accuracy between lineups. We included a practice session, as the results from Experiment 1 suggested that this improved both accuracy with the actual lineup and the reliability of the pupillary data. Finally, we asked participants to type the number of the face they were looking at, so that we could be sure that they had seen the number.

Experiment 2

Methods

Participants

In the target-present condition, 66 participants were recruited in exchange for cash or university course credits. Four were removed due to technical issues, leaving sixty-two participants (9 males and 53 females) aged between eighteen and thirty-five ($M = 20.60$, $SD = 3.19$). Participants were recruited until there were at least ten for each category of

Measuring Recognition in Lineups with Pupillometry

identification response. This resulted in 36 identifiers, 14 non-identifiers and 12 misidentifiers in each lineup. In the target-absent condition there were 22 participants (5 males, 16 females, and one participant who did not provide their gender), aged between 18 and 40 ($M = 22.19$, $SD = 6.04$). This resulted in 10 misidentifiers and 12 correct rejectors.

Apparatus and Materials

These were the same as in the previous experiment, except that nine lineup videos were used, each assigned a number from 1-9. The number was displayed clearly at the top left of the screen throughout each video clip. In the target-absent lineup, participants only saw the eight distractors that had been presented in the target-present display.

Design

This was the same as the previous experiment.

Procedure

This was also the same as the previous experiment, except that if participants thought they recognised one of the faces as that of the perpetrator, they needed to remember its allocated number, as they were asked for this at the end of the lineup.

Results

Target-Present condition.

We conducted a two-way mixed ANOVA to determine whether pupil sizes changed in response to the target (see fig. 4). For the first lineup, there was a significant effect of

Measuring Recognition in Lineups with Pupillometry

identification response, $F(2,59) = 4.97$, $p = .01$, $r = .28$, $\eta^2 = .20$ (Identifiers: $M = 61.30$, $SE = 2.10$; Misidentifiers: $M = 50.50$, $SE = 3.60$; Non-identifiers: $M = 51.70$, $SE = 3.40$). There was no significant effect of *face type*, $F(2,118) = 2.98$, $p = .06$ (Pre: $M = 51.90$, $SE = 2.70$; Target: $M = 59.40$, $SE = 3.00$; Post: $M = 52.20$, $SE = 2.30$) and no significant interaction between *face type* and *identification response*, $F(4,118) = 1.92$, $p = .11$.

[Insert Fig. 4 here]

Planned contrasts revealed that pupil sizes for *identifiers* were significantly larger for the target than for pre-target distractors, $F(1,35) = 12.72$, $p = .01$, $r = .52$ (14%) and post-target distractors, $F(1,35) = 35.18$, $p < .001$, $r = .71$ (18%). For *non-identifiers*, pupils were no larger for the target than for pre-target distractors $F(1,13) = 0.01$, $p = .97$ (0.3%), or post-target distractors, $F(1,13) = 0.01$, $p = .98$ (-0.2%). There were also no significant comparisons for *misidentifiers*: pre-target distractors, $F(1,11) = 2.08$, $p = .18$ (8%), post-target distractors $F(1,11) = 0.01$, $p = .93$ (0.8%).

Between identifiers, non-identifiers and misidentifiers, three one-way ANOVAs revealed significant pupillary differences for the target, $F(2,61) = 7.24$, $p = .01$, $r = .33$, but not for pre-target distractors, $F(2,61) = 2.41$, $p = .10$, or post-target distractors, $F(2,61) = 0.01$, $p = .96$.

As in Experiment 1, we then conducted a paired samples *t*-test for each identification response, to compare pupil sizes when people looked at the target compared to when looking at the LD. For all three identification responses pupil sizes were significantly larger when looking at the LD compared to when looking at the target: *misidentifiers*, $t(11) = 3.98$, $p = .01$; *non-identifiers*, $t(13) = 4.00$, $p = .01$; *identifiers*, $t(35) = 5.11$, $p = .01$.

Measuring Recognition in Lineups with Pupillometry

[Insert Table 3 here]

For the second lineup, there were no significant effects of *face type*, $F(2,118) = 2.62$, $p = .08$ (Pre: $M = 45.95$, $SE = 2.52$; Target: $M = 50.63$, $SE = 3.25$; Post: $M = 43.79$, $SE = 2.81$) or *identification response*, $F(2,59) = 1.06$, $p = .35$ (Identifiers: $M = 47.99$, $SE = 2.51$; Misidentifiers: $M = 41.80$, $SE = 4.47$; Non-identifiers: $M = 50.83$, $SE = 4.74$). However, there was a significant interaction between *face type* and *identification response*, $F(4,118) = 5.90$, $p < .001$, $\eta^2 = .17$.

Planned contrasts revealed that pupil sizes for *identifiers* were significantly larger for the target than for pre-target distractors, $F(1,37) = 37.61$, $p < .001$, $r = .71$ (19%) and post-target distractors $F(1,37) = 29.36$, $p < .001$, $r = .67$ (19%). For *non-identifiers*, pupils were not significantly larger for the target than for pre-target distractors $F(1,11) = 2.72$, $p = .13$ (-5%), or post-target distractors, $F(1,11) = 1.11$, $p = .31$ (7%). There were also no significant comparisons for *misidentifiers*: pre-target distractors, $F(1,11) = 0.01$, $p = .96$ (-0.3%), post-target distractors $F(1,11) = 0.79$, $p = .39$ (-6%).

Between identifiers, non-identifiers and misidentifiers, three one-way ANOVAs revealed significant pupillary differences for the target, $F(2,61) = 4.34$, $p = .02$, $r = .26$, for pre-target distractors, $F(2,61) = 3.77$, $p = .03$, $r = .24$, but not for post-target distractors, $F(2,61) = 0.24$, $p = .79$.

Three *t*-tests again revealed that for all three identification responses pupil sizes were significantly larger when looking at the LD compared to when looking at the target: *misidentifiers*, $t(11) = 5.56$, $p = .01$; *non-identifiers*, $t(11) = 4.82$, $p = .01$; *identifiers*, $t(37) = 3.40$, $p = .01$.

Measuring Recognition in Lineups with Pupillometry

[Insert Table 4 here]

Thus, pupillary responses separated identifiers, misidentifiers and non-identifiers and separated face types in both lineups (albeit in slightly different ways).

Using pupil size to predict identification response.

As in Experiment 1, two binary logistic regressions were used to determine whether pupil size change predicted whether participants made a correct or incorrect lineup decision. For the first lineup, the model was statistically significant, $\chi^2(1) = 7.34, p = .01$, explaining 15% (Nagelkerke R^2) of the variance in lineup decision outcome and correctly classifying 67.7% of cases. For the second lineup, the logistic regression model was also statistically significant, $\chi^2(1) = 11.16, p = .01$. The model explained 22.4% (Nagelkerke R^2) of the variance in lineup decision outcome and correctly classified 72.6% of cases. Therefore, pupil size was a fairly good measure of identification performance in both lineups.

Target-Absent condition

Two two-way mixed ANOVAs were again used for analysis of pupil size measures in response to each TA lineup. As can be seen in fig 5, there were no significant effects for either lineup. For the first lineup: *face type*, $F(2,40) = 0.30, p = .75$ (Pre: $M = 57.00, SE = 3.80$; Target: $M = 52.60, SE = 5.90$; Post: $M = 55.20, SE = 3.80$); *identification response*, $F(1,20) = 2.58, p = .12$ (Misidentifiers: $M = 60.00, SE = 4.70$; Correct rejectors: $M = 49.80, SE = 4.30$). There was also no significant interaction between *face type* and *identification response*, $F(2,40) = 1.07, p = .35$.

[Insert Fig. 5 here]

Measuring Recognition in Lineups with Pupillometry

Inspection of fig. 5 appears to show that pupillary responses for the target were different for misidentifiers and correct rejectors, but a one-way ANOVA confirmed that this was not the case, $F(1,21) = 2.78, p = .11$. Thus, pupillary responses did not discriminate significantly between misidentifiers and correct rejectors or between face types in either lineup.

Again, we conducted a paired samples t -test for each identification response, to compare pupil sizes when people looked at the false positive compared to when looking at the LF. For both misidentifiers and correct rejectors, pupil sizes were significantly larger when looking at the LF compared to when looking at the false positive: *Misidentifiers*: $t(9) = 2.57, p = .03$; *Correct rejectors*: $t(11) = 4.74, p = .03$.

[Insert Table 5 here]

For the second lineup, (fig. 5): *face type*, $F(2,40) = 1.41, p = .25$ (Pre: $M = 45.70, SE = 4.40$; Target: $M = 45.90, SE = 5.00$; Post: $M = 38.60, SE = 4.20$); *identification response*, $F(1,20) = 0.37, p = .55$ (Misidentifiers: $M = 45.60, SE = 5.00$; Correct rejectors: $M = 41.20, SE = 5.00$). There was also no significant interaction between *face type* and *identification response*, $F(2,40) = 0.99, p = .38$. As there were no significant effects, we did not conduct t -tests.

Using pupil size to predict identification response:

Again, two binary logistic regressions were used to determine whether pupil size change predicted whether participants made a correct or incorrect lineup decision. Neither

Measuring Recognition in Lineups with Pupillometry

regression was significant, $\chi^2(1) = 2.44$, $p = .12$; second lineup, $\chi^2(1) = 3.14$, $p = .08$. Pupil size did not predict whether participants would correctly reject or misidentify.

Discussion

Experiment two investigated whether pupillometry would benefit current UK police procedures, where eyewitnesses memorise the number of a face in a lineup if they wish to identify it. We predicted that participants who recognised the target in the target-present condition would have larger pupils when looking at his face compared to the distractors' faces. We also predicted that there would be no significant pupillary changes in the target-absent condition. Both predictions were confirmed, but in experiment two, unlike experiment one, the pupillary changes in the target-present condition occurred with both lineups.

An explanation for this effect may lie in competing accounts of pupillary responses to cognitive processing. Experiment two, like experiment one, was based on the assumption that pupillary changes were associated with memory strength, as pupils have been shown to be larger when memory strength is greater (Brocher & Graf, 2016; Otero et al., 2011; Papesh et al., 2012; review in Goldinger & Papesh, 2012). In experiment one, this assumption was supported by the data, as pupillary changes were only found for the first target-present lineup (when *only* the target face had been seen before), and were greatest in participants who both claimed to remember his face *and* made an identification.

In experiment two, in the first TP lineup, pupillary changes *only* occurred in participants who made an identification. Therefore, pupillary changes differentiated between participants who had identified the target and those who had not. In the second TP lineup, pupillary changes differentiated between all three identification response groups: pupils again

Measuring Recognition in Lineups with Pupillometry

only responded to the target in participants who identified him; they gradually became slightly smaller in people who made no identification; and they did not change in size in participants who misidentified a distractor.

In this version of the procedure, pupillary changes were observed in participants who identified the target as they looked at him in both lineups. Why this occurred is not clear. In the second TP lineup, all the faces would now have been familiar to the participants. Therefore, if pupil size simply responds to memory, the faces in the second TP lineup should all have elicited similar pupil sizes, but this was not the case. A memory strength account is nevertheless still plausible, as the target's face would have been the most familiar to those people who recognised him (as they had been exposed to his face more). Therefore, the pupil sizes of those who recognised him would have been somewhat larger when looking at his face compared with the less-familiar distractors. However, this account alone fails to explain why similar effects did not occur with the second TP lineup presentation in experiment one. Alternative theories for the pupillary effects will be discussed in the general discussion.

We also investigated whether pupil size change could predict whether participants made a correct or incorrect lineup decision, and found that pupil sizes successfully predicted participants' decisions in both lineups of the target-present condition. This is important, as finding a way of measuring an individual eyewitness's pupillary data, and comparing them to a model for predicting correct decisions could be useful to police, particularly when making assessments about the credibility of an eyewitness's identification response.

However, to test real world systems, it is best to do so in the real world. Therefore, for Experiment 3, we tested pupillary lineups with a portable eye-tracker, using members of the

Measuring Recognition in Lineups with Pupillometry

general public. We were also interested in the effects of anxiety in witnesses, as anxiety is common during crimes and lineups (Miller & Bornstein, 2013), but is difficult to investigate in laboratory experiments. We took inspiration from Valentine and Mesout (2009), who found that only 17% of those who were anxious when they encountered a frightening actor in the London Dungeon identified him in a subsequent lineup, compared to 75% of those who were not anxious. We focused on evaluating the effects of anxiety on *pupillary responses*. Larger pupil sizes are associated with stress, fear and trauma (Bitsios, Szabadi, & Bradshaw, 1996; Kimble, Fleming, Bandy, Kim, & Zambetti, 2010). Therefore, we considered it to be important to test whether pupillometry would be effective to measure implicit recognition in anxious participants.

Experiment 3

Methods

Participants

Twenty participants (12 males and 8 females) aged between eighteen and fifty-three ($M = 27.35$, $SD = 8.14$) were recruited from the Booster Ride on Brighton Pier during the British Science Festival (2017).

Apparatus and Materials

In 2017, the British Science Festival was based in Brighton and had an evening event, where different experiments, presentations and events were set up. We conducted our experiment at this event, taking advantage of Brighton Pier's Booster Ride. This is a scary

Measuring Recognition in Lineups with Pupillometry

five-minute ride that launches two pairs of people high into the air and swings them back down at high speeds.

Experiment Builder was run on a 21.5 inch laptop computer and a portable EyeLink Duo eye-tracker, which uses an infrared camera to provide precise measures of gaze location and pupil size.

The stimuli consisted of 8 colour video clips of white female faces (aged between 25-30) against a grey background. The video clips were constructed as they had been in the previous experiments, according to the VIPER criteria.

A questionnaire consisting of 20 multiple choice questions was used to ascertain the emotional state of participants while they were on the ride, and one question asked if they had been anxious or not. The questionnaire took approximately two minutes to complete.

Design

This was the same as in the previous experiments.

Procedure

People waiting to ride on the Booster Ride were recruited by a female researcher. After the ride, those who had agreed to take part in the experiment were asked by her to complete a short questionnaire designed to evaluate their level of anxiety while on the ride. She spent approximately 2-3 minutes with each participant before they proceeded to the lineup, but was near the ride throughout the event.

Measuring Recognition in Lineups with Pupillometry

After completing the questionnaire, each participant was directed to a nearby tent, where a pupillary lineup was conducted by a second researcher. The participant did not see the first researcher again. At this point, they placed their chin on the chin rest, and their eye movements were calibrated to five points on the computer screen. After reading instructions on the screen, their gaze was monitored with a drift check, which involved looking at a black dot on a white screen. They were told that they would be presented with a lineup that might include the face of the first researcher. They then saw a video lineup that included her face and seven distractors. The rest of the procedure was similar to the previous experiments, except that the practice session was omitted due to time constraints, and it was only possible to use a target-present lineup.

After data collection, participants were grouped according to self-reported anxiety. This was based on the final question in the questionnaire, which asked them to rate their level of anxiety while on the ride, on a scale from 1-5 (1 = not anxious at all, and 5 = very anxious). One participant failed to record their level of anxiety, so the responses of the other 19 participants were analysed. 12 of these rated their level of anxiety as 1 (not anxious at all), and 7 chose a rating that was greater than 1, indicating that they experienced at least some anxiety. Therefore, we decided to divide participants into two anxiety groups: "anxious" (7 participants) and "non-anxious" (12 participants).

Results

Again, we used two two-way mixed ANOVAs. Inspection of fig. 6. reveals that for the first lineup, there was an effect of *identification response*, $F(2,17) = 4.22$, $p = .03$, $r = .45$, $\eta^2 = .33$ (Identifiers: $M = 65.38$, $SE = 4.87$; Misidentifiers: $M = 43.40$, $SE = 6.89$; Non-

Measuring Recognition in Lineups with Pupillometry

identifiers: $M = 47.83$, $SE = 6.89$), but there was no significant effect of *face type*, $F(2,34) = 2.30$, $p = .12$ (Pre: $M = 48.58$, $SE = 5.52$; Target: $M = 59.35$, $SE = 4.85$; Post: $M = 48.68$, $SE = 4.33$). There was also no significant interaction between *face type* and *identification response*, $F(4,34) = 2.53$, $p = .06$, $\eta^2 = .23$.

[Insert Fig. 6 here]

Planned contrasts revealed that pupil sizes for *identifiers* were significantly larger for the target than for pre-target distractors, $F(1,9) = 9.74$, $p = .01$, $r = .72$, $\eta^2 = .52$ (28%) and post-target distractors $F(1,9) = 28.17$, $p < .001$, $r = .87$, $\eta^2 = .76$ (32%). For *non-identifiers*, pupils were not significantly larger for the target than for post-target distractors, $F(1,4) = 0.06$, $p = .82$, $\eta^2 = .01$ (-2%). They were also not significantly larger for the pre-target distractors $F(1,4) = 0.02$, $p = .89$ (2%). There were also no significant comparisons for *misidentifiers*: pre-target distractors, $F(1,4) = 0.57$, $p = .49$ (7%), or post-target distractors $F(1,4) = 0.06$, $p = .83$ (-2%).

Between identifiers, non-identifiers and misidentifiers, three one-way ANOVAs showed that there were significant pupillary differences for the target, $F(2,19) = 8.88$, $p = .01$, $r = .56$, but not for pre-target distractors, $F(2,19) = 1.15$, $p = .34$, or post-target distractors, $F(2,19) = 0.37$, $p = .70$.

We conducted a paired samples *t*-test for each identification response, to see whether pupil sizes were different when people looked at the target compared to when looking at the LD. For *misidentifiers*, pupil sizes were significantly larger when looking at the LD compared to when looking at the target, $t(4) = 3.31$, $p = .03$. For *non-identifiers*, pupil sizes

Measuring Recognition in Lineups with Pupillometry

were not significantly different between the two face types, $t(4) = 2.26, p = .09$. For *identifiers*, pupil sizes were also not significantly different between the face types, $t(9) = 0.02, p = .99$.

[Insert Table 6 here]

Also presented in fig. 6, for the second lineup, there were no significant effects: *face type*, $F(2,34) = 2.52, p = .10$ (Pre: $M = 53.81, SE = 5.59$; Target: $M = 44.42, SE = 6.33$; Post: $M = 41.54, SE = 4.65$); *identification response*, $F(2,34) = 0.12, p = .89$ (Identifiers: $M = 46.83, SE = 5.12$; Misidentifiers: $M = 43.76, SE = 7.59$; Non-identifiers: $M = 49.19, SE = 8.48$). There was also no significant interaction between *face type* and *identification response*, $F(4,34) = 0.88, p = .49$. As there were no significant effects, we did not conduct *t*-tests as we had done for the first lineup.

Thus, pupillary responses were no different between participants when they were divided according to their identification response, or between the different face types.

Using pupil size to predict identification response.

Two binary logistic regressions were used to determine whether pupil size change predicted whether participants made a correct or incorrect lineup decision. For the first lineup, the model was statistically significant, $\chi^2(1) = 8.50, p = .01$, explaining 46.2% (Nagelkerke R^2) of the variance in lineup decision outcome and correctly classifying 75% of cases. For the first lineup, pupil size was therefore a good measure of identification performance. This was not true for the second lineup, for which the logistic regression was not significant, $\chi^2(1) = 3.15, p = .08$.

Measuring Recognition in Lineups with Pupillometry

Pupil sizes in anxious and non-anxious participants:

Anxiety Groups

A two-way mixed ANOVA was performed to see whether pupil sizes were different when people with different levels of anxiety looked at the faces in both lineups. There was one within-subjects factor, *lineup presentation* (with two levels: First lineup and Second lineup) and one between-subjects factor: *level of anxiety* (with two levels: Anxious and Non-anxious).

There was a significant effect of *anxiety*, $F(1, 17) = 8.97, p = .01, r = .59$ (Anxious: $M = 57.86, SE = 3.69$; Non-anxious: $M = 43.95, SE = 2.82$): pupil sizes of anxious participants were larger than those of non-anxious participants. There was no significant difference in pupil size between the first lineup and the second, $F(1, 17) = 3.79, p = .07$, and no interaction between *lineup presentation* and *anxiety*, $F(1, 17) = 0.01, p = .93$. Thus, the pupils of anxious participants were larger than those who were not anxious throughout the experiment.

Associations between pupillary changes and level of anxiety:

Pupillary changes

To see whether pupillary changes for the different faces were different between anxious and non-anxious participants, we calculated the mean difference in pupil size for *all* participants when viewing the target and distractors. In the first lineup, this difference was 15% (Target: $M = 66.80, SE = 6.22$; Distractors: $M = 51.14, SE = 3.89$), $t(19) = 2.88, p = .01, r = .50$. We therefore categorised all participants with a pupillary change of 15% or more

Measuring Recognition in Lineups with Pupillometry

when viewing the target as “pupillary changers” and those with a pupillary change of less than 15% when viewing the target as “pupillary non-changers”.

Fisher’s Exact tests were then performed on level of anxiety (anxious and non-anxious) and pupillary changes (pupillary changers and pupillary non-changers). For the first lineup, there was no significant association between anxiety group and pupillary response, $\chi^2(1) = 0.09, p = .57$ (Non-anxious: 50% = pupillary change, 50% = no pupillary change; Anxious: 43% = pupillary change, 57% = no pupillary change). Thus, while the pupil sizes of anxious participants were larger than those of non-anxious participants, this did not affect their responses to the faces. This analysis was not conducted on the second lineup, as there was no difference in pupil size between the two face types.

Discussion

Our previous results were replicated in a field experiment: participants who identified the target had larger pupils when viewing her face than when viewing the faces of distractors, but only when viewing the lineup for the first time. Interestingly, anxiety seems to have made no difference to these fluctuations in pupillary responses associated with recognition, implying that pupillometry could be an appropriate identification measure to use with anxious people. However, these results need to be replicated with a larger sample than used in the present experiments. Finally, we wanted to test the use of a portable eye-tracker for forensic purposes, and found that it was ideal. The EyeLink Duo (SR Research, 2017) was easy to use and set up, processed participants quickly, and provided precise pupillary data. It demonstrated that a portable system would be a valid addition to police procedures, by being

Measuring Recognition in Lineups with Pupillometry

practical for use in the homes of witnesses and victims, thus reducing the anxiety associated with police lineups (Miller & Bornstein, 2013).

Poor weather conditions meant that we could test only 20 participants. Although this limited our analyses, our results were consistent with previous research in showing that pupil sizes were larger in anxious participants (Kimble et al., 2010). Numerous procedural differences preclude any close comparisons between our study and Valentine & Mesout's (2009) work on the effects of anxiety on identification. For one thing, they tested recognition of the face of the person who had been the source of the anxiety, whereas we investigated the effect of anxiety on recognising the face of someone who had not caused the anxiety. This is important, as con-artists commit crimes without frightening their victims, and in some frightening crimes, it can be difficult to know who the perpetrator was.

General Discussion

All three experiments demonstrated that pupil sizes changed in response to a target face, but only in participants who correctly identified it, and generally only the first time they saw it. The only exception was in Experiment 2, when pupils also responded to the target in the second lineup. In all three experiments, we were also able to *predict* identification of the targets from participants' pupillary responses in the first lineup. However, when the target was absent, pupillary responses did not differentiate between correct rejectors and misidentifiers. Thus, it was the presence of a previously-seen face that resulted in the pupillary effects in the target-present lineups. This approach is in line with research that indicates that pupillary changes are associated with memory strength (Brocher & Graf, 2016;

Measuring Recognition in Lineups with Pupillometry

Otero et al., 2011; Papesh et al., 2012), and we concluded that the pupillary responses probably reflected implicit memory strength for the target's face. This will be discussed further below.

We wanted to assess pupillometry in relation to traditional eyewitness identification responses, which rely on conscious processes and have been shown to be unreliable (Wells, Steblay, & Dysart, 2015). One could gain more insight into the thought processes underlying the witness' decision, for instance, by asking them if they were Remembering, Knowing or Guessing. However, these would still be conscious inferences made by the witness about their own decision-making processes. By bypassing the witness' explicit evaluation of their decision, pupillometry appears to overcome these issues. If an eyewitness makes an identification, but their pupils do not respond to the identified face, they are likely to be wrong. If they make no identification, but their pupils get larger when seeing the suspect, they may well have recognised them implicitly.

Our findings have relevance for police lineup procedures. Previous research has suggested that sequential lineups reduce false identification rates compared to simultaneous lineups; however, this advantage disappears if the sequential lineup is shown twice, as in the U.K. system (review in Steblay et al., 2011). Our findings provide further evidence that there are no advantages to providing eyewitnesses with a second lineup presentation. In Experiments 1 and 3, there were no pupillary changes to the target in identifiers in the second lineups. Changes were detected in experiment 2, but the second lineup did not elicit any meaningful new effects that had not already occurred in the first.

Measuring Recognition in Lineups with Pupillometry

On a more theoretical level, the pupillary findings can be evaluated in terms of three constructs: cognitive load, cognitive engagement, and memory strength.

Cognitive load is the notion that task demands affect mental workload (Sweller, 2010). Pupil sizes are larger when cognitive load is high and smaller when cognitive load is low (Chen & Epps, 2014; Jainta & Baccino, 2010; Piquado, Isaacowitz, & Wingfield, 2010; Zekveld et al., 2014). Processing faces that have only been seen briefly before is likely to place a greater cognitive burden on limited working memory resources than processing familiar faces, so cognitive load theory would predict that pupil sizes should be *smaller* when looking at familiar faces. Our finding that pupil sizes were *larger* when viewing the more familiar targets (in participants who identified them) offers no support for an explanation in terms of cognitive load – although, of course, this is not a strong argument, as the target face would only be weakly familiar in absolute terms.

Another explanation is in terms of cognitive engagement. There is evidence that pupil sizes are larger when looking at salient objects (that have emotional content or are socially important) than non-salient objects (e.g. Bradley et al. 2008; Laeng & Falkenberg, 2007; Partala & Surakka, 2003; Prehn et al., 2011; Snowden et al., 2016; Võ et al., 2008). Given that faces are socially important, it is likely that they will also differ in terms of engagement: pupils should be larger when looking at a socially-important face than a socially-unimportant one. It is also likely that a previously-seen face will be more socially-important than a novel one, so the pupils should be larger when viewing the previously-seen targets than the novel distractors. In a lineup paradigm, these pupillary responses might well be amplified, as larger pupils are also associated with goal-seeking (Mathôt, Siebold, Donk, & Vitu, 2015) and reward (Satterthwaite et al., 2007). Given the pressure to perform well in lineups (Miller &

Measuring Recognition in Lineups with Pupillometry

Bornstein, 2013), it is likely that participants would be motivated to do well, and this would be reflected in their pupil sizes, particularly when looking at a target that appears familiar. Our findings are consistent with an account in terms of cognitive engagement. As pointed out by one of our reviewers, an alternative explanation, that fits in with engagement theory, is that participants try to keep in mind the face selected in the first lineup when looking at the second lineup presentation. This might explain why Experiment 2 produced pupil changes in both lineup presentations.

Finally, it is possible to interpret our findings in terms of memory strength. This is determined by the similarity between the item held in memory and the target item, and decisions are based on strength of evidence (Dunn, 2008). Pupils are larger when memory strength is greater (Brocher & Graf, 2016; Goldinger & Papesh, 2012; Otero et al., 2011; Papesh et al., 2012). They may also reflect the experience of recognition (Otero et al., 2011), and the strength of evidence on which recognition is based (Montefinese et al., 2018). Therefore, faces that are more familiar should elicit a larger pupil size when they are recognised than weakly-represented (totally novel) faces. Our findings are consistent with this account. Overall, however, it has to be concluded that they do not strongly favour one theoretical explanation over another. Not only do the memory strength and cognitive engagement explanations make similar predictions about pupil size, but both phenomena might conceivably be operating in tandem to produce the observed effects.

This research lends itself to a variety of modifications that could clarify other issues related to police lineup identifications. Systems such as VIPER enable witnesses to conduct lineup procedures independently by following on-screen instructions. However, as pointed out by one of our reviewers, it would be interesting to see whether pupil sizes might be

Measuring Recognition in Lineups with Pupillometry

affected by feedback (e.g. confirming or disconfirming a witness's lineup decision) or response bias (such as police lineup instructions). Other useful research could be to test the pupillary responses of less communicative individuals in police lineups (e.g. children), who may find the procedure less intimidating without having to make a conscious decision. It would also be worth testing with people with face processing deficits, to investigate whether their pupils could respond to a face in the absence of a correct identification decision.

Empirically, it seems clear that pupils respond differently to faces that differ in terms of familiarity, and probably reflect the different implicit mental processes involved. The findings tentatively suggest that the pupillary responses to the target might reflect memory strength or perhaps cognitive engagement, rather than cognitive load. However further evaluation and analysis of possible theoretical accounts should be explored in future research: as one of our reviewers pointed out, it is important that identification techniques have a sound theoretical basis. In conjunction with the use of implicit and explicit measures of performance, lineup paradigms are a promising technique for doing this. We do not propose that pupillary responses can replace explicit responses in police lineups, but our research suggests they can offer a measure of implicit mental processing that provides insight into the identification responses that people make. Consequently, pupillometry appears to highlight the issues with current police lineup procedures, and could be used to support them.

References

- Ayres, P., & Paas, F. (2012). Cognitive Load Theory: New directions and challenges. *Applied Cognitive Psychology*, 26(6), 827–832. <https://doi.org/10.1002/acp.2882>

Measuring Recognition in Lineups with Pupillometry

- Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, 91(2), 276-292.
- Bindemann, M., Brown, C., Koyas, T., & Russ, A. (2012). Individual differences in face identification predict eyewitness accuracy. *Journal of Applied Research in Memory and Cognition*, 1(2), 96–103. <https://doi.org/10.1016/j.jarmac.2012.02.001>
- Bitsios, P., Szabadi, E., & Bradshaw, C. M. (1996). The inhibition of the pupillary light reflex by the threat of an electric shock: a potential laboratory model of human anxiety. *Journal of Psychopharmacology*, 10(4), 279–287.
- Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology*, 45(4), 602–607. <https://doi.org/10.1111/j.1469-8986.2008.00654.x>
- Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior*, 26(3), 353-364.
- Brewer, N., & Palmer, M. A. (2010). Eyewitness identification tests. *Legal and Criminological Psychology*, 15(1), 77–96. <https://doi.org/10.1348/135532509X414765>
- Brocher, A., & Graf, T. (2016). Pupil old/new effects reflect stimulus encoding and decoding in short-term memory. *Psychophysiology*, 53(12), 1823–1835. <https://doi.org/10.1111/psyp.12770>

Measuring Recognition in Lineups with Pupillometry

- Bruce, V., Henderson, Z., Newman, C., & Burton, A. M. (2001). Matching identities of familiar and unfamiliar faces caught on CCTV images. *Journal of Experimental Psychology: Applied*, 7(3), 207–218. <https://doi.org/10.1037//1076-898X.7.3.207>
- Burton, A. M., Jenkins, R., & Schweinberger, S. R. (2011). Mental representations of familiar faces. *British Journal of Psychology*, 102(4), 943–958.
<https://doi.org/10.1111/j.2044-8295.2011.02039.x>
- Burton, A. M., White, D., & McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, 42(1), 286–291.
- Chen, S., & Epps, J. (2014). Using task-induced pupil diameter and blink rate to infer cognitive load. *Human–Computer Interaction*, 29(4), 390–413.
<https://doi.org/10.1080/07370024.2014.892428>
- Conway, M. A., & Dewhurst, S. A. (1995). Remembering, familiarity, and source monitoring. *The Quarterly Journal of Experimental Psychology Section A*, 48(1), 125–140. <https://doi.org/10.1080/14640749508401380>
- Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior*, 12(1), 41–56.
- Dunn, J. C. (2004). Remember-Know: A matter of confidence. *Psychological Review*, 111(2), 524–542. <https://doi.org/10.1037/0033-295X.111.2.524>
- Dwyer, J., Neufeld, P., & Scheck, B. (2000). *Actual innocence: five days to execution and other dispatches from the wrongly convicted* (1st ed). New York: Doubleday.

Measuring Recognition in Lineups with Pupillometry

Goldinger, S. D., & Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Current Directions in Psychological Science*, 21(2), 90–95.

<https://doi.org/10.1177/0963721412436811>

Goldinger, S. D., He, Y., & Papesh, M. H. (2009). Deficits in cross-race face learning: Insights from eye movements and pupillometry. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(5), 1105–1122.

<https://doi.org/10.1037/a0016548>

Hancock, P. J., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences*, 4(9), 330–337.

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>

The Innocence Project (n.d.), retrieved 18th May, 2018, from

<https://www.innocenceproject.org/causes/eyewitness-misidentification/>

Jainta, S., & Baccino, T. (2010). Analyzing the pupil response due to increased cognitive demand: An independent component analysis study. *International Journal of Psychophysiology*, 77(1), 1–7. <https://doi.org/10.1016/j.ijpsycho.2010.03.008>

Kimble, M. O., Fleming, K., Bandy, C., Kim, J., & Zambetti, A. (2010). Eye tracking and visual attention to threatening stimuli in veterans of the Iraq war. *Journal of Anxiety Disorders*, 24(3), 293–299. <https://doi.org/10.1016/j.janxdis.2009.12.006>

Measuring Recognition in Lineups with Pupillometry

Körner, C., Braunstein, V., Stangl, M., Schlögl, A., Neuper, C., & Ischebeck, A. (2014).

Sequential effects in continued visual search: Using fixation-related potentials to compare distractor processing before and after target detection. *Psychophysiology*, 51(4), 385-395.

Hsiao, J. H. W., & Cottrell, G. (2008). Two fixations suffice in face recognition. *Psychological Science*, 19(10), 998-1006.

Laeng, B., & Falkenberg, L. (2007). Women's pupillary responses to sexually significant others during the hormonal cycle. *Hormones and Behavior*, 52(4), 520–530.
<https://doi.org/10.1016/j.yhbeh.2007.07.013>

Lefebvre, C. D., Marchand, Y., Smith, S. M., & Connolly, J. F. (2007). Determining eyewitness identification accuracy using event-related brain potentials (ERPs). *Psychophysiology*, 44(6), 894–904. <https://doi.org/10.1111/j.1469-8986.2007.00566.x>

MacLin, O. H., MacLin, M. K., & Malpass, R. S. (2001). Race, arousal, attention, exposure and delay: An examination of factors moderating face recognition. *Psychology, Public Policy, and Law*, 7(1), 134–152. <https://doi.org/10.1037//1076-8971.7.1.134>

Mansour, J. K., & Flowe, H. D. (2010). Eye tracking and eyewitness memory. © British Psychological Society.

Mathôt, S., Siebold, A., Donk, M., & Vitu, F. (2015). Large pupils predict goal-driven eye movements. *Journal of Experimental Psychology: General*, 144(3), 513–521.
<https://doi.org/10.1037/a0039168>

Measuring Recognition in Lineups with Pupillometry

Miller, M. K., & Bornstein, B. H. (Eds.). (2013). *Stress, Trauma, and Wellbeing in the Legal System*. Oxford, UK: Oxford University Press. DOI:

10.1093/acprof:oso/9780199829996.001.0001

Montefinese, M., Vinson, D., & Ambrosini, E. (2018). Recognition memory and featural similarity between concepts: the pupil's point of view. *Biological psychology*, 135, 159-169.

Murphy, G., Groeger, J. A., & Greene, C. M. (2016). Twenty years of load theory—Where are we now, and where should we go next? *Psychonomic Bulletin & Review*, 23(5), 1316-1340. <https://doi.org/10.3758/s13423-015-0982-5>

My Light Meter Pro (2015), retrieved from, <https://itunes.apple.com/gb/app/mylightmeter-pro/id583922375?mt=8>, 18th August 2017.

National Research Council. (2015). *Identifying the culprit: Assessing eyewitness identification*. National Academies Press.

Orzechowski, J., Nęcka, E., & Balas, R. (2016). Task conditions and short-term memory search: two-phase model of STM search. *Polish Psychological Bulletin*, 47(1), 12-20.

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory: Pupil size and recognition memory. *Psychophysiology*, 48(10), 1346–1353. <https://doi.org/10.1111/j.1469-8986.2011.01217.x>

Measuring Recognition in Lineups with Pupillometry

- Papesh, M. H., Goldinger, S. D., & Hout, M. C. (2012). Memory strength and specificity revealed by pupillometry. *International Journal of Psychophysiology*, 83(1), 56–64.
<https://doi.org/10.1016/j.ijpsycho.2011.10.002>
- Partala, T., & Surakka, V. (2003). Pupil size variation as an indication of affective processing. *International Journal of Human-Computer Studies*, 59(1–2), 185–198.
[https://doi.org/10.1016/S1071-5819\(03\)00017-X](https://doi.org/10.1016/S1071-5819(03)00017-X)
- Peters, D. P. (1988). Eyewitness memory and arousal in a natural setting. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life* (pp. 89–94). Chichester, UK: Wiley.
- Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, 47(3), 560–569.
<https://doi.org/10.1111/j.1469-8986.2009.00947.x>
- Prehn, K., Heekeren, H. R., & van der Meer, E. (2011). Influence of affective significance on different levels of processing using pupil dilation in an analogical reasoning task. *International Journal of Psychophysiology*, 79(2), 236–243.
<https://doi.org/10.1016/j.ijpsycho.2010.10.014>
- Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: Evidence from pupillometry and event-related fMRI. *NeuroImage*, 37(3), 1017–1031. <https://doi.org/10.1016/j.neuroimage.2007.04.066>

Measuring Recognition in Lineups with Pupillometry

Sauer, J., Brewer, N., Zweck, T., & Weber, N. (2009). The effect of retention interval on the Confidence–Accuracy relationship for eyewitness identification. *Law and Human Behavior*, 34(4), 337–347. <https://doi.org/10.1007/s10979-009-9192-x>

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied*, 15(1), 46–62. <https://doi.org/10.1037/a0014560>

Seale-Carlisle, T. M., & Mickes, L. (2016). US line-ups outperform UK line-ups. *Royal Society Open Science*, 3(9), 160300. <https://doi.org/10.1098/rsos.160300>

Snowden, R. J., O’Farrell, K. R., Burley, D., Erichsen, J. T., Newton, N. V., & Gray, N. S. (2016). The pupil’s response to affective pictures: Role of image duration, habituation, and viewing mode. *Psychophysiology*, 53(8), 1217–1223. <https://doi.org/10.1111/psyp.12668>

SR Research (n.d.), retrieved, April 23rd, 2018, from <https://www.sr-research.com/products/eyelink-1000-plus/>

Stebay, N. K., Dietrich, H. L., Ryan, S. L., Raczynski, J. L., & James, K. A. (2011). Sequential lineup laps and eyewitness accuracy. *Law and Human Behavior*, 35(4), 262–274. <https://doi.org/10.1007/s10979-010-9236-2>

Stebay, N., Dysart, J., Fulero, S., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior*, 25(5), 459–473. <https://doi.org/10.1023/A:1012888715007>

Measuring Recognition in Lineups with Pupillometry

Sweller, J. (2010). Cognitive Load Theory: Recent theoretical advances. In J. L. Plass, R.

Moreno, & R. Brunken (Eds.), *Cognitive Load Theory* (pp. 29–47).

Valentine, T., & Mesout, J. (2009). Eyewitness identification under stress in the London

Dungeon. *Applied Cognitive Psychology*, 23(2), 151–161.

<https://doi.org/10.1002/acp.1463>

van Rijn, H., Dalenberg, J. R., Borst, J. P., & Sprenger, S. A. (2012). Pupil dilation co-varies

with memory strength of individual traces in a delayed response paired-associate task.

PLoS ONE, 7(12), e51134. <https://doi.org/10.1371/journal.pone.0051134>

VIPER (n.d.), retrieved January 16th, 2018, from <http://www.viper.police.uk>

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>

Wells, G. L., & Bradfield, A. L. (1998). " Good, you identified the suspect": Feedback to

eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied*

Psychology, 83(3), 360-376.

Wells, G. L., Ferguson, T. J., & Lindsay, R. C. (1981). The tractability of eyewitness

confidence and its implications for triers of fact. *Journal of Applied Psychology*,

66(6), 688-696.

Measuring Recognition in Lineups with Pupillometry

Wells, G. L., Small, M., Penrod, S. J., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E.

(1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, 22, 603-647.

Wells, G. L., & Olson, E. A. (2003). Eyewitness Testimony. *Annual Review of Psychology*, 54(1), 277–295. <https://doi.org/10.1146/annurev.psych.54.101601.145028>

Wells, G. L., Steblay, N. K., & Dysart, J. E. (2015). Double-blind photo lineups using actual eyewitnesses: An experimental test of a sequential versus simultaneous lineup procedure. *Law and Human Behavior*, 39(1), 1–14.
<https://doi.org/10.1037/lhb0000096>

Wixted, J. T., Read, D. J., & Lindsay, S. D. (2016). The effect of retention interval on the eyewitness identification Confidence–Accuracy relationship. *Journal of Applied Research in Memory and Cognition*, 5(2), 192–203.
<https://doi.org/10.1016/j.jarmac.2016.04.006>

Wright, D. B., & Stroud, J. N. (2002). Age differences in lineup identification accuracy: people are better with their own age. *Law and Human Behavior*, 26(6), 641.

Zekveld, A. A., Heslenfeld, D. J., Johnsrude, I. S., Versfeld, N. J., & Kramer, S. E. (2014). The eye as a window to the listening brain: Neural correlates of pupil size as a measure of cognitive listening load. *NeuroImage*, 101, 76–86.
<https://doi.org/10.1016/j.neuroimage.2014.06.069>