Encoding-linked pupil response is modulated by expected and unexpected novelty: Implications for memory formation and neurotransmission

Alex Kafkas

University of Manchester, School of Biological Sciences, Division of Neuroscience and Experimental Psychology, Manchester, UK

Abstract

Whether a novel stimulus is expected or unexpected may have implications for the kind of ensuing encoding and the type of subsequent memory. Pupil response was used in the present study to explore the way expected and unexpected stimuli are encoded and whether encoding-linked pupil response is modulated by expectation. Participants first established a contingency relationship between a series of symbols and the type of stimulus (man-made or natural) that followed each one. At encoding, some of the target stimuli violated the previously established relationship (i.e., unexpected), while the majority conformed to this relationship (i.e., expected). Expectation at encoding had opposite effects on familiarity and recollection, the two types of memory that support recognition, and modulated differently the way pupil responses predicted subsequent memory. Encoding of unexpected novel stimuli was associated with increased pupil dilation as a predictor of subsequent memory type and strength. In contrast, encoding of expected novel stimuli was associated with decreased pupil response (constriction), which was predictive of subsequent memory type and strength. The findings support the close link between pupil response and memory formation, but critically indicate that this is modulated by the type of novelty as defined by expectation. These novel findings have important implications for the encoding mechanisms involved when different types of novelty are detected and is proposed to indicate the operation of different neurotransmitters in memory formation.

Keywords: expectation, novelty, pupil dilation, pupil constriction, recognition memory, encoding

Correspondence:

Dr Alex Kafkas

University of Manchester

Oxford Road

Zochonis Building

Manchester, M13 9PL

Email: alexandros.kafkas@manchester.ac.uk

Tel.: +44 (0) 161 275 7341

"There is nothing permanent except change"
Heraclitus (c.535 – 475BC)

1. Introduction

Accurate prediction of the sequence of events may be seen as an attractive alternative to the uncertainty we so often face in everyday life. The human mind strives to form predictions about regularities in the environment (Friston et al., 2015), but this effort is disrupted when unexpected encounters occur that violate previous monotones. However, expectedness, when new stimuli are detected and encoded into memory, can have consequences for the mechanisms that support learning and memory (Kafkas & Montaldi, 2015a, 2018a). Besides, expectation and especially unexpectedness is closely related to novelty (Kafkas & Montaldi, 2018b; Reichardt, Polner, & Simor, 2020) and the experience of surprise (Barto, Mirolli, & Baldassarre, 2013). Understanding the way expectation affects encoding of new stimuli is still lacking, but its exploration can benefit theories of learning and memory. Therefore, the interplay between expectation, novelty type and encoding of information is the focus of the present study.

As novelty can have diverse meanings (for reviews see Barto et al., 2013; Frank & Kafkas, 2020; Kafkas & Montaldi, 2018b; Ranganath & Rainer, 2003), in the present study, the term is used to refer to any experience or stimulus which lacks specific pre-existing representations from an episodic standpoint. In this sense, novelty may be based on features of the item itself (e.g., the picture of dice has not been presented before in the experiment, therefore the item is new) or on features of the context a stimulus is presented in (e.g., the picture of dice has not been presented, but its presentation is also surprising and unexpected in the current context). Finally, conceptual or semantic novelty, where a stimulus completely lacks pre-experimental conceptual representations as it happens, for example, in the case of abstract shapes or non-words, will not be examined here.

1.1. How does expectation define novelty and modulate memory encoding and retrieval?

Novelty detection itself is motivational (Bunzeck & Duzel, 2006) as a novel bit of information may indicate something significant in the environment that we may want to remember and integrate with existing knowledge, either to avoid it in the future or to reap possible rewards. However, the concept of novelty is not univocal in that different types of novelty can be discriminated each of which may have implications for the way we process, encode, act upon and remember them in the future (Frank & Kafkas, 2020; Kafkas & Montaldi, 2018b; Schomaker & Meeter, 2015). Novelty may enhance episodic memory formation directly as novel stimuli are prioritised at encoding (e.g., Geraci & Manzano, 2010; Tulving & Kroll, 1995) or indirectly as encoding benefits may ensue any stimulus presented in the context of novelty (e.g., Bunzeck & Duzel, 2006; Fenker et al., 2008; Kafkas & Montaldi, 2015a; Okuda, Højgaard, Privitera, Bayraktar, & Takeuchi, 2020; Schomaker, van Bronkhorst, & Meeter, 2014). However, the beneficial effect of novelty on memory is not absolute in that it does not always lead to subsequent memory boosts (e.g., Biel & Bunzeck, 2019). This indicates that novelty is prioritised in memory under certain conditions but not always.

Expectation may be an important dimension that allows different types of novelty to be discriminated. For example, an unexpected novel stimulus may trigger more rigorous encoding mechanisms (or attentional resource allocation) than an expected novel stimulus. However, the direction of episodic memory effects defined by the encoding of expected or unexpected stimuli is far from straightforward. For example, previous studies have revealed that stimuli conforming to previously established conceptual associations tend to be remembered better relative to incongruent stimuli (the congruency effect; e.g., Bartlett, 1932; Bein et al., 2015; Craik & Tulving, 1975; Staresina, Gray, & Davachi, 2008). In

contrast, distinctive information that deviates from the perceptual or conceptual characteristics of the current context, and is therefore unexpected, has also been shown to result in enhanced memory encoding as revealed by greater subsequent memory retrieval (e.g., Rangel-Gomez & Meeter, 2013; Schmidt & Schmidt, 2017; Wallace, 1965). Furthermore, two recent studies (Greve, Cooper, Tibon, & Henson, 2019; Kafkas & Montaldi, 2018a), using expectation manipulations related to violation or validation of newly established rules, have found memory advantage for both expected and unexpected stimuli.

Greve et al., (2019) showed that a memory advantage can be seen not only for stimuli that are consistent with experimentally established arbitrary associations, but also for stimuli that violated this association (i.e., unexpected). Kafkas & Montaldi (2018a) showed that stimuli that conform to learned contingency rules (expected) and those that violate these rule (unexpected) are accompanied by increased memory performance based, nevertheless, on different types of memory. Expected stimuli were remembered better on the basis of familiarity, while unexpected stimuli were accompanied by increased recollection accuracy. Familiarity denotes memory for the item itself, without recall of associative details from the previous occurrence, whereas recollection indicates retrieval of associative information from the time of encoding (e.g., Kafkas & Migo, 2009; Kafkas & Montaldi, 2012; Mandler, 1980; Montaldi & Mayes, 2010; Yonelinas, 2002). Therefore, processing of expected and unexpected novel stimuli may implicate different encoding mechanisms, which they lead to qualitatively different memory experiences. The main aim in the present study was to explore whether encoding mechanisms triggered from the detection of expected and unexpected stimuli vary. To this end pupil response measures were employed during encoding and their relationship to subsequent type of memory were monitored adopting a subsequent memory effect protocol (Paller & Wagner, 2002).

1.2. Encoding-linked pupil responses and modulation by type of novelty

Before examining the specific hypotheses of this study, it is relevant to consider the evidence for the close link between episodic memory encoding and pupil response and what implications this link may have for encoding mechanisms. It has been shown that the pupil response at encoding can be predictive of the type and strength of ensuing memory (e.g., Kafkas & Montaldi, 2011, 2015a; Miller & Unsworth, 2019; Naber et al., 2013; Papesh, Goldinger, & Hout, 2012; Wetzel, Einhäuser, & Widmann, 2020). However, the studies appear inconsistent as to the pattern of pupil response that predicts subsequent memory. For example, increased pupil dilation for later remembered relative to forgotten stimuli has been reported (e.g., Miller & Unsworth, 2019; Papesh et al., 2012), but this effect appears to be more pronounced when contextual or unexpected novelty is detected (Kafkas & Montaldi, 2015a). On the other hand, other studies have shown that decreased pupil dilation or pupil constriction predict the type and/or the strength of subsequent memory (Kafkas & Montaldi, 2011; Naber et al., 2013; Wetzel et al., 2020). This finding is also consistent with the pupil old/new effect, whereby reduced levels of pupil dilation accompany new relative to old stimuli in recognition memory paradigms (e.g., Kafkas & Montaldi, 2012, 2015b; Otero, Weekes, & Hutton, 2011; Vö et al., 2008).

In an effort to account for the obvious inconsistency in the way pupil response reflects successful memory formation for novel stimuli, Kafkas and Montaldi (2018b), in a recent neurocognitive model, proposed that different neural pathways and neurotransmitter systems are tuned in to different types of novelty. Specifically, contextual, surprising and unexpected novelty modulates the noradrenergic sympathetic system resulting in dopamine and norepinephrine release in the hippocampus and increased pupil dilation. On the other hand, according to the same model, detection of absolute and expected novelty modulates the cholinergic parasympathetic system resulting in acetylcholine release in the hippocampus and pupil constriction patterns. Similar to the proposal by Kafkas & Montaldi

(2018b), recent reviews (Joshi & Gold, 2020; Peinkhofer, Knudsen, Moretti, & Kondziella, 2019) have also highlighted the utility of pupil response patterns in understanding brain pathways involved in cognition.

1.3. The present study

In the present study, the encoding mechanisms engaged when expected and unexpected stimuli are encountered were explored and their effect on subsequent memory were measured. The pupil response was used as a measure of online processing during the encoding period. Overall, the experiment aimed at exploring how expectation modulates the encoding mechanisms engaged for successful (i.e., subsequent familiarity or recollection hits) and unsuccessful (i.e., a subsequently forgotten/missed items) memory formation. It is assumed that in the case of expected stimuli, only stimulus absolute novelty is in operation at encoding as the sequence of the event conform to the established rule, while a new stimulus is presented. On the other hand, in the case of unexpected stimuli, apart from stimulus absolute novelty, a degree of contextual novelty and surprise is in operation due to the violation of expected sequence of the event. Therefore, all stimuli used at encoding were experimentally novel (i.e., objects presented for the first time in the experiment), but a subset of them also violated contextual expectation.

The following specific hypotheses were explored: Based on findings from (Kafkas & Montaldi, 2018a), it was hypothesised that expected and unexpected stimuli will result in opposite effect on subsequent familiarity and recollection responses. Based on the memory-related pupil effects reported in previous studies (see 1.3.), it was hypothesised that pupil response patterns at encoding will be predictive of the type and possibly the strength of reported memory. However, consistent with the proposal from Kafkas and Montaldi (2018b), it was hypothesised that these pupil response patterns at encoding will be modulated by the expectation status of the stimuli resulting perhaps in different constriction or dilation patterns for expected and unexpected stimuli. Finally, whether the hypothesised pupil modulation by expectation only applies to subsequent recollections or it also affects subsequent familiarity responses, was also explored.

2. Material and Methods

2.1. Participants

A total of 40 participants gave informed consent and participated in the study, which received ethical approval from the University of Manchester Research Ethics Committee. From this original sample, 3 participants were excluded from all analyses due to incomplete/noisy eye tracking recordings (see 2.4 below; inclusion of their behavioural data did not change the reported memory effects). The final sample included 37 participants (34 females, 3 males) with a mean age of 19.32 years (SD = 2.75). The sample size was informed by previous studies with similar methodology (Kafkas & Montaldi, 2011, 2018a). Specifically, a power analysis using GPower tool (Faul, Erdfelder, Lang, & Buchner, 2007), determined that 36 participants is a sufficient sample to replicate the previously reported (Kafkas & Montaldi, 2018a) effect of expectation on memory (parameters: $\eta^2 = 0.11$; effect size $f = 0.35$; power = 0.90). All participants had normal or corrected-to-normal (with contact lenses) vision and self-reported no history of psychiatric or neurological conditions. They were further asked to abstain from coffee and alcohol consumption 24 hours before participating in the study, while systematic use of psychotropic medicines and drugs were described as exclusion criteria.

2.2. Stimuli

A subset of stimuli from Kafkas and Montaldi (2018a) were used in the present study. Briefly, 246 simple man-made and natural items presented in grayscale were used in the

different conditions of the experiment (size: 500 x 375 pixels; degrees of visual angle at presentation: 18.7h and 14.05v). As has been described in detail before (see Kafkas & Montaldi, 2011, 2015b), the stimuli were matched for luminosity and emitted luminance and were presented in grayscale, with low contrast on a uniform grey background (RGB = 130; Figure 1). These steps ensured that the stimuli are appropriate for pupillometry. The encoding and recognition tasks included 170 stimuli (85 man-made and 85 natural), while 72 stimuli (36 man-made and 36 natural) were used in the rule learning task (see 2.3.). Four stimuli were also used for practice. Finally, 6 symbols (line drawings) were used in the rule learning task and as contextual cues at encoding (see 2.3. and Figure 1).

2.3. Procedure and design

Each experimental session consisted of 4 tasks: the rule learning, the encoding, the filled interval and the recognition task (Figure 1). The design of the experiment was similar to the design reported in Experiment 1a in Kafkas and Montaldi (2018a). The only difference was that eye tracking data were recorded at encoding and additional interstimulus interval was implemented to ensure recovery of pupil size from trial to trial (see below). The allocation of the object stimuli across the different conditions (rule learning, encoding, recognition) was completely random with the only restriction being the type of stimulus (man-made or natural) consistent with the expectation manipulation.

Figure 1. Experimental design. a) The rule learning task established contingency symbol-stimulus type associations, with each symbol (see d) consistently associated with either a man-made or a natural item (3 symbols assigned to each type). Twelve repetitions of the same symbol across the entire block were used. b) At encoding, a novel set of stimuli (completely different from those used in a) were presented after the same cuing symbols from the rule learning task. Sixty percent (60%) of the trials adhered to the established (in the rule learning task) rule (i.e., expected stimuli), while 40% of trials violated the previously learned rule (i.e., unexpected stimuli). The sequence within an expected and an unexpected event are illustrated in the examples. In the case of the unexpected stimulus (top sequence), the stimulus (padlock)

is inconsistent with the rule established previously (i.e., symbol should be followed by a natural item). In the case of the expected stimulus (bottom sequence), the stimulus (dice) is consistent with the rule established previously (i.e., symbol should be followed by a man-made item). c) A recognition memory task tested memory for the previously encoded expected and unexpected stimuli intermixed with new (unstudied) foils. Participants pressed the corresponding button, while each stimulus was on the screen to report stimuli as either new (N), familiar (F) or recollected (R). d) All 6 symbols used as cuing stimuli in the experiment.

2.3.1. The rule learning task

In the rule learning task participants learned a contingency relationship between a symbol and the type (man-made or natural) of the subsequent item. In each trial participants were asked to make a prediction as to whether a natural or a man-made object will follow a symbol (see Figure 1a). In total 6 symbols (Figure 1d) were used across this block and each of them was randomly assigned (across participants) to precede either a man-made (3 symbols) or a natural (3 symbols) item. The association remained the same throughout the rule learning task. In each trial the symbol remained on the screen until participants responded using two keyboard buttons to predict either a man-made or a natural item. After the participants' prediction, a blank gray screen was presented for 1s followed by the object stimulus presented for 3s. Each symbol was repeated 12 times across the entire task, preceding a unique stimulus in each trial, which was nevertheless consistent with the specific symbol-stimulus type rule.

As shown in Figure 1a, for the first 36 trials (6 repetitions per symbol) a feedback screen followed the presentation of the stimulus, indicating whether the symbol-related prediction was correct or incorrect. For the next 18 trials no feedback was provided. Finally, the last 18 trials involved the same symbol-stimulus sequence, but participants were not asked to make a prediction. In these last trials the symbol remained on the screen for 2s. The sequence within the last 18 trials made the transition to the encoding task smoother. Overall, the design of this block ensured effective learning of the contingency relationships and the build-up of expectations based on the correct symbol-stimulus sequences (Kafkas & Montaldi, 2018a).

2.3.2. The encoding task

Before participants started the encoding task, a standard nine-point calibration of the eye tracking system was performed, while each participant's head was stabilised using an appropriate chinrest. Participants were informed that their eye movements would be recorded during this block to measure attention on the presented stimuli, but no explicit reference to pupil responses or to the type of memory test that would follow was made. This task was very similar to the final part of the rule learning task (last 18 trials; see 2.3.1) in that a symbol-stimulus sequence was adopted. The main difference was that in the encoding task participants were asked to focus on studying the stimuli carefully without the need to make any prediction in relation to the symbols. The same 6 symbols as in the rule learning task were used (20 trials each symbol), but a completely new set of man-made/natural stimuli (120 in total) were presented. The critical manipulation in this block was that only 60% of the trials (72 trials), spread across the entire encoding block, adhered to the previously established symbol-stimulus type relationship (expected stimuli). The remaining 40% of trials violated the symbol-stimulus rule and they were, therefore, unexpected. In the encoding task participants were asked to freely view the presented stimulus, without performing any other encoding task.

Each encoding trial started with a fixation cross (1s), followed by a blank gray screen, which served as a baseline period (1s) for calculating the trial-specific pupil response. The cuing symbol appeared for 2s followed by a blank screen (1s) and then the stimulus that

remained on the screen for 3s. To ensure recovery of stimulus-related change in pupil size, a delay period of 2s was used after each stimulus and was marked with a gray fixation cross. Participants had the chance to take 2 self-paced breaks (every 40 trials) during this task.

2.3.3. The recognition memory task

After completing the encoding task, each participant engaged in a filled interval of 10 minutes during which unrelated verbal and arithmetic tasks were performed. Subsequently, training and instructions pertaining the recognition memory decisions were offered. Occurrences of familiarity (F), recollection (R) and novelty (N) were explained and participants had the chance to practice these choices with 4 practice stimuli, two of which had been presented previously as practice items just before the encoding task. It was explained that a recollection response indicated the recall of additional contextual details associated with the encoding episode (e.g., a thought related to a specific stimulus). Contrary, a familiarity response meant that no additional detail came to mind related to the study episode apart from memory for the item itself. Similar to previous work with fMRI and eye tracking methodologies (e.g., Kafkas, Mayes, & Montaldi, 2020; Kafkas et al., 2017; Kafkas & Montaldi, 2012, 2014; Mayes et al., 2019) familiarity responses were provided using a 3-point rating scale from weak to strong familiarity. Participants explicitly instructed not to confuse confidence with the familiarity and recollection decision. Instead, it was stressed that they can be very confident they have encountered a stimulus before at encoding without being able to recollect additional details associated with this encounter (for previous work related to this important distinction see e.g., Kafkas et al., 2017; Kafkas & Montaldi, 2012; Mayes et al., 2019).

Before commencing the main part of this task, participants had the chance to ask questions in relation to the recognition decisions and they were also asked to explain to the experimenter what the response choices meant. This ensured appropriate use of the recognition memory decisions and accurate reporting of the experienced memory kind for each stimulus. Participants completed a total of 170 trials (120 from encoding and 50 new ones), each starting with a fixation cross. Each stimulus appeared for a maximum duration of 3s, during which participants were instructed to provide their recognition decision. This timing was also practiced in the practice trials before commencing the main block. No eye tracking data were recorded during this period and therefore a chinrest was not used.

2.3.4. Debriefing

After the end of the experimental session participants were asked to explain what they thought the specific aim of the study was and were subsequently informed of the aims by the experimenter. As in the previous study in which a similar paradigm was used (Kafkas & Montaldi, 2018a), only a small number of participants (2) reported they realised the violation of the symbol-stimulus rule in some of the encoding trials. The two participants were included in all the analyses, but exclusion of their data did not change the direction of the main findings as reported here.

2.4. Apparatus, eye tracking recording and pre-processing

Apparatus: Each session was completed in a quiet testing room moderately lit (at about 250 lx) across all sessions and equipped with an ASL infrared eye tracking system (Applied Science Laboratories, Model Eye-Trac 6000; sampling rate 60 Hz). A standard-nine-point calibration was performed with each participant just before stating the encoding task. An artificial 4mm pupil was used at the end of encoding at the position of each participant's eye to later convert the pupil size recordings into mm. All parts of the experiment were completed using a PC with a 19-inch monitor (resolution: 1280 x 1024). Participants had a distance of about 70cm from the computer monitor and a special chinrest was using for the period of eye tracking recording (encoding task). A standard computer keyboard was used to collect participants' behavioural responses.

Pre-processing of the pupil data followed previously established protocols (e.g., Kafkas & Montaldi, 2011, 2015b). Specifically, blinks, artefacts and partial closures of the eyelid were identified on the pupil size recordings by the eye tracking software and were excluded from the analyses. Also, to ensure that no pupil trace was contaminated by other residual artefacts and noise, a grand mean was calculated across all pupil traces within each trial and pupil recordings diverging more than three standard deviations from the mean were also discarded. These traces are artefacts of the pupil recording mostly related to partial eyelid closures right before or after a blink. Trials with more than 50% of discarded traces were excluded from analyses and participants with too many excluded trials (more than 40% of total number) were removed from the sample (3 participants were removed; see 2.1.) As the discarded pupil recordings across all trials, in the final participant sample, were very low (less than about 3% of the total recorded traces), no interpolation was employed to replace these recordings. As in previous studies, (see e.g., Kafkas & Montaldi, 2015b), linear interpolation for the missing traces did not have any effect on the observed pupil effects and therefore analyses with non-interpolated data are reported below.

Pupil timeseries: The trial-related pupil recording lasted for 3s, which was the duration of each trial at encoding, giving therefore a total of 180 continuous pupil recordings per trial (if no blinks or artefacts were identified during this period). The average pupil size during a period of 1s preceding each symbol was used as trial-specific baseline. Pupil timeseries for each trial were calculated by subtracting the trial-specific baseline average from each one of the 180 pupil size recordings during the presentation of the target stimulus. Therefore, the pupil timeseries were expressed as deviations from a zero-point baseline allowing the discrimination of pupil dilation and constriction patterns. It is relevant to stress that the terms pupil dilation and constriction relate to variations in pupil size due to the presence of a stimulus relative to a preceding baseline period. They should not be considered as indicating an absolute match onto the physiological dilation and constriction responses of the pupil. That said, a reduction in pupil response is more likely to evince physiological constriction of the pupillary sphincter, while increased pupil response is more likely to relate to physiological dilation of the iris dilator muscle. Finally, for the analyses, the pupil timeseries were divided in 10 timebins of 300ms. These timeseries were the main pupil response measures reported in this experiment¹.

¹Average or peak pupil responses can be used to summarise the trial-related pupil changes but their utility is limited as they do not allow evaluation of the time course of pupil change during the information processing period.

2.5. Statistical analyses

Linear mixed effect analyses using maximum likelihood estimation were performed to analyse the effect of expectation and memory type on memory performance, response times at recognition (RTs) and pupil response at encoding. Memory performance was calculated using d' (d -prime) separately for familiarity and recollection responses. These analyses were performed using the GAMLj module (Gallucci, 2019) in Jamovi 1.2.16 (The jamovi project, 2020), which implements linear mixed effect analysis based on the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R software. An important advantage of modelling predictors (independent variables) and measures (dependent variables) using linear mixed effects, instead of analysing them using repeated measures ANOVAs, include the ability to control for the variance attributed to random factors, such as the participants in the sample or any other factor that may explain random variance across participants (Baayen, Davidson, & Bates, 2008). It is also better suited for unbalanced designs where the number of observations or available data may vary across participants or conditions (as in the present study).

In the linear mixed effects model for memory performance, expectation (expected versus unexpected) and subsequent memory type (familiar versus recollected) were set as fixed effect factors (predictors), while participants were specified as a random effect (random intercept). The same modelling approach was also adopted for the analysis of RTs. Similarly, to analyse the effect of expectation and memory type on encoding-related pupil response, expectation, memory type and time window at encoding were set as fixed factors, while the random factor of participants on the intercept was also fitted. The time variable included the pupil response over the period of encoding (3s) summarised into 10 timebins (moving window of 300ms). The memory type in this analysis included missed (M), familiar (F) and recollected (R) stimuli. For each analysis, type III Wald F tests (with Satterthwaite's method for approximating degrees of freedom; Kuznetsova, Brockhoff, & Christensen, 2017; Satterthwaite, 1946) are reported and significant effects were further explored using pairwise comparisons with Bonferroni adjustment for multiple comparisons. For the main analyses, familiarity responses were collapsed across the three levels of rating, but they were also examined separately in the case of pupil response to explore the relationship between familiarity strength and pupil response.

To determine which random effects to include in the analyses, models with random effects of memory type and expectation status across participants (i.e., random slope of conditions across subjects) were fitted for comparison. These are summarised in Table 1. Including random slopes in mixed effects models has been shown to reduce type I and type II errors (Schieleth & Forstmeier, 2009). The model that significantly improved the explained variance using fewer parameters included the memory type as random effect across participants for performance and RTs and both memory type and expectation as random effects for pupil response. The findings from these models are presented in the Results and the comparison parameters of the different hierarchical models from each analysis are summarised in Tables A.1-A.3 (Appendix A). For these comparisons, the likelihood ratio tests approach was adopted, using the χ^2 distribution, which compares whether the addition of a parameter (in this case a random effect) significantly improves the fit of the model. Additionally, changes in the variance explained by the models when fixed and random effects are included were also taken into account using a pseudo R^2 measure (Johnson, 2014; Nakagawa & Schielzeth, 2013). Specifically, R^2_{marginal} indicates the variance explained by the fixed effects only relative to the total variance of the dependent measure, while $R^2_{\text{conditional}}$ indicates the variance explained by the fixed and random effects together relative to the variance in the dependent measure. These metrics are also provided in the description of the results as indirect indicators of effect size. A significance level of $p < 0.05$ was adopted for all analyses.

Table 1. Three nested models with the same fixed factors (expectation, memory type and expectation by memory type interaction) but increased complexity in the included random factor for the analysis of the dependent variables (DV)

Model (using lme4 call formatting in R)	Model explanation
1) DV ~ 1 + expectation + memory type + expectation:memory type +(1 subjects)	Explores the effects of fixed factors on DV with random intercept for participants
2) DV ~ 1 + expectation + memory type + expectation:memory type +(1 + memory type subjects)	Explores the effects of fixed factors on DV with random slope of memory type across participants
3) DV ~ 1 + expectation + memory type + expectation:memory type +(1 + memory type + expectation subjects)	Explores the effects of fixed factors on DV with random slope of memory type and expectation across participants

3. Results

3.1. Behavioural effects

3.1.1. Rule learning

Mean accuracy across participants and symbol repetition in the rule learning task was 0.86 (SD = 0.12). Importantly, as shown in Figure 2, symbol prediction accuracy significantly increased with repeated presentations of the symbols ($F_{11,396} = 27.6, p < 0.001$), while RTs were significantly reduced across symbols repetitions ($F_{11,396} = 7.86, p < 0.001$). Therefore, the rule learning task was successful at creating clear expectations regarding the symbol-stimulus sequences.

Figure 2. Prediction accuracy and response times (RTs) across 12 repetitions of each symbol in the rule learning task. Significant improvement in accuracy and reduction in RTs indicate successful learning of the contingency associations and therefore development of clear expectations for symbol-stimulus sequences.

3.1.1 Differential effect of expectation on memory performance for familiarity and recollection

A linear mixed effect analysis was conducted to explore the effect of expectation status at encoding (expected versus unexpected) and type of subsequent memory (familiar versus recollected) on memory performance (measured using d'). As can be seen in Table A.1, the model that included the random effect of participants (intercept) and the random effect of memory type significantly improved the data fit over the simpler (fixed effects + random effect of participants) or the more complex model (fixed effects + random effects of participants, memory type and expectation). This model explained large proportion of variance in memory performance ($R^2_{\text{marginal}} = 0.30$; $R^2_{\text{conditional}} = 0.77$) and significantly improved the fit. This means that the magnitude of the effect of memory type on memory performance varies across participants and therefore their random slopes should be modelled in the analysis of fixed effects presented below.

Table 2. Mean proportions and RTs for the different response outcomes at recognition for expected and unexpected (at encoding) stimuli

Response	Proportions		RTs (in ms)	
	Expected	Unexpected	Expected	Unexpected
F_{hits}	0.82(0.09)	0.67(0.13)	1354(201)	1362 (210)
R_{hits}	0.10(0.05)	0.20(0.06)	1359(300)	1341 (296)
M	0.12(0.08)	0.20(0.10)	1427(316)	1335 (329)
F_{FA}	0.28(0.16)		1404(281)	1401 (271)
R_{FA}	0.01(0.03)		1425(368)	1379 (508)
CR	0.72(0.16)		1272(217)	1272 (217)

Note: Numbers in the parentheses are standard deviations. F_{hits} =familiarity hits; R_{hits} = recollection hits; M = misses; FA = false alarms; F_{FA} = familiarity false alarms; R_{FA} = recollection false alarms; CR= correct rejections. R_{hits} are means across 22 participants who reported recollections, while the rest of the proportions are means across 37 participants².

The mean proportions across the different response outcomes (hits, false alarms, misses and correct rejections) for expected and unexpected stimuli at encoding are summarised in Table 2. The analysis showed that the expectation status of stimuli at encoding did not predict overall subsequent memory performance ($F_{1,74} = 0.91$, $p = 0.34$), however, significantly higher performance associated with familiarity responses relative to recollection responses ($F_{1,37} = 24.62$, $p < 0.001$, $\beta = -0.75$, $SE = 0.12$). Importantly, the expectation by memory type interaction was significant ($F_{1,74} = 44.49$, $p < 0.001$) indicating that differences in memory performance for familiarity and recollection responses varied across expectation status at encoding ($\beta = 0.83$, $SE = 0.12$). Specifically, post hoc tests showed that familiar responses to expected stimuli were associated with increased memory performance relative to familiar responses to unexpected stimuli ($t_{76.1} = 5.32$, $p_{\text{bonferroni}} < 0.001$). Opposite to this effect, recollection responses to unexpected stimuli were associated with increased memory performance relative to recollections for expected stimuli ($t_{76.1} = -3.99$, $p_{\text{bonferroni}} < 0.001$; Figure 3). Therefore, the findings indicate that the expectation status at encoding has opposite effects on memory performance for familiarity and recollection responses, with the latter favouring unexpected stimuli and the former

² This is the reason the proportions of old stimuli ($F_{\text{hits}} + R_{\text{hits}} + M$) do not add-up to 1.00 as R_{hits} are calculated for $N = 22$ (participants reporting Rs), while F_{hits} and M are calculated for $N = 37$.

favouring expected stimuli. Similar effects were found when d' was calculated using the criterion of independence for familiarity responses (Yonelinas & Jacoby, 1995) or when using the proportion of hits minus the proportion of false alarms as a measure of memory performance.

Figure 3. Novel expected and unexpected stimuli at encoding have opposite effects on familiarity and recollection responses at retrieval. a) Memory performance (d') for subsequent familiarity and recollection responses for stimuli encoded as expected or unexpected. Error bars show standard error of the mean. b) Distribution of subsequent familiarity and recollection responses for expected and unexpected stimuli at encoding. Percentages represent division of the total number of recollection (upper panel) and total number of familiarity responses (lower panel) across expected and unexpected stimuli.

3.1.2. RTs analysis

RTs were analysed in the same way as memory performance. Inclusion of memory type as a random effect significantly improved the model fit (for the different metrics and comparisons across the 3 models; see Table A.2), resulting in increased $R^2_{\text{conditional}}$ of 0.84. The R^2_{marginal} of 0.008 shows that the fixed factors do not reliably predict the variance in RTs. Indeed, neither the main effect of expectation ($F_{1,53.3} < 1$) nor the main effect of response ($F_{1,23.6} = 1.00$, $p = 0.32$) nor the expectation by response interaction ($F_{1,53.9} = 1.24$, $p = 0.27$) were significant. Therefore, the expectation status at encoding and the type of memory at recognition did not have any effect on the RTs at recognition.

3.2. Pupil response at encoding by stimulus expectation and subsequent memory type

A linear mixed effect analysis was conducted to explore the effect of expectation status at encoding (expected versus unexpected), memory type (M, F, R) and time (encoding duration divided in 10 time-bins of 300ms) on pupil response. As can be presented in Table A.3, the model that included the random effect of participants (intercept), memory type and expectation (i.e., random slope of conditions of interest on pupil response) significantly improved the data fit ($R^2_{\text{marginal}} = 0.24$; $R^2_{\text{conditional}} = 0.46$). This means that the magnitude of the effect of memory type and expectation on pupil response varies across participants and therefore their random slopes should be modelled in the analysis of fixed effects presented below.

The analysis showed that the pupil response at encoding was significantly higher for unexpected relative to expected stimuli ($F_{1,39.8} = 52.35, p < 0.001; \beta = 0.06, SE = 0.008$), but it did not differ significantly across the three memory types ($F_{2,29} = 2.03, p = 0.150$). However, the memory type at retrieval significantly predicted the degree of pupil response at encoding, but the shape of this effect was different depending on the expectation status of stimuli at encoding. This finding was qualified by the significant expectation by memory type interaction ($F_{2,1791} = 121.54, p < 0.001$) and the significant triple interaction between expectation, memory type and time ($F_{18,1742} = 11.18, p < 0.001$). As shown in Figure 4, for expected stimuli there was a significant linear reduction of pupil response during encoding as a function of subsequent memory type (across M, F, R; $\beta_{linear} = -0.05, SE = 0.009, p < 0.001$). In contrast, for the unexpected stimuli, a significant linear increase of pupil response during encoding was associated with later memory type (across M, F, R; $\beta_{linear} = 0.08, SE = 0.009, p < 0.001$). This differential linear trend of the pupil response by the subsequent memory type, depending on the expectation status of stimuli, became significant half-way through the encoding period around 1500ms after stimulus onset for both expected ($\beta = -0.04, SE = 0.081, p = 0.03$) and unexpected stimuli ($\beta = 0.04, SE = 0.019, p = 0.02$), and remained significant until the end of the encoding period (i.e., 3s). Similar differential linear effects were also found when examining subsequent memory type separated by familiarity strength (i.e., M, F1, F2, F3, R) for expected ($\beta = -0.05, SE = 0.009, p < 0.01$) and unexpected stimuli ($\beta = 0.08, SE = 0.009, p < 0.01$). The only difference in the effect was that the linear increase in pupil response across memory type became significant slightly later (1800ms) in the case of unexpected stimuli.

4. Discussion

In the present study, pupil response was examined during the encoding of expected and unexpected novel stimuli. The results replicated the previously reported finding (Kafkas & Montaldi, 2018a) that the encounter of expected and unexpected stimuli have contrasting effect on familiarity and recollection, the two types of memory that support recognition memory. Importantly, it was shown that the pupil response patterns at encoding are modulated by the expectedness of the encoded stimuli. Specifically, expected stimuli that conform to previously established regularities, were accompanied by a linear reduction (or constriction) of pupil response tracking the type and the strength of subsequent memory. On the other hand, encoding of unexpected stimuli that disrupt contextual or temporal regularities in the sequence of events, resulted in increased pupil response patterns (or dilations) predicting the strength and the type of subsequent memory.

This mirror-image pattern following the detection of expected and unexpected novel stimuli is consistent with the hypothesis that distinct encoding mechanisms are involved when the expectedness of novelty varies (for this proposal see Kafkas & Montaldi, 2018b). These expectation-modulated encoding mechanisms, as revealed by the pupil response patterns, were not only relevant for recollection, but also affected memory formation resulting in familiarity. The implications of these findings for novelty encoding, memory performance and the utility of pupil response as indicator of memory formation are discussed below. Also, in line with proposals that pupil response patterns provide a window into the involved neural circuitry (Joshi & Gold, 2020; Kafkas & Montaldi, 2018b) and the known relationship between sympathetic and parasympathetic pathways in controlling dilation and constriction patterns (Steinhauer & Hakerem, 1992; Steinhauer, Siegle, Condray, & Pless, 2004; see also Widmann, Schröger, & Wetzels, 2018), the involvement of different neurotransmitter systems in encoding expected and unexpected novelty are also discussed.

Figure 4. Pupil response at encoding across expected (a) and unexpected stimuli (b) as a function of subsequent memory type and familiarity strength. M = missed old stimuli; F = familiarity hits collapsed and separately across F1 (weak), F2 (moderate) and F3 (strong) familiarity; R = recollected hits. The degree of pupil response at encoding significantly predicted subsequent type of memory (or memory strength), but with opposite directions according to the expectation status of the encoded stimuli. The period during which the linear effect of subsequent memory on pupil response is significant ($p < 0.05$) is marked.

4.1. Expectation, memory formation and implications for type of memory

The behavioural findings show that whether a stimulus is expected or unexpected has consequences for the type of memory that is formed, which is reflected subsequently when memory is probed. Expected stimuli, that conform to established rules, are more likely to be remembered on the basis of familiarity, whereas unexpected stimuli, that violate the established rule, are more likely to produce accurate recollections. This opposite effect of expectation on the kind of memory supporting recognition is comparable to the previously reported effect using a similar paradigm (Kafkas & Montaldi, 2018a). However, in the

previous paradigm, apart from the expectation status of stimuli, the type of encoding also varied. In the present study, the effect of expectation was replicated without requiring different encoding tasks, confirming, therefore, that the differential effect of expectation on later memory type cannot be explained by decision variations at encoding.

To the extent that unexpected stimuli are more distinctive at encoding, we may assume that these would receive preferential attentional allocation (Schmidt & Schmidt, 2017). This suggestion agrees with the general pattern of pupil response that accompanied unexpected stimuli at encoding. Traditionally, increased pupil dilations are thought to indicate enhanced resource allocation, enhanced processing and effort (e.g., Granholm, Asarnow, Sarkin, & Dykes, 1996; Hess & Polt, 1964; Kahneman, 1973; see discussion of pupil patterns below in 4.2.). But is this selective attentional allocation to the distinctive unexpected items at the expense of expected stimuli in terms of subsequent memory performance? Based on the present findings, the answer has to be twofold as it depends on the type of memory that is formed and supports retrieval. When examining recollection alone, there is a clear advantage for unexpected stimuli, therefore a presumed selective attentional allocation at encoding appears to be advantageous for the distinctive (unexpected) items. Consistent with this, recent neurocomputational evidence (Frank, Montemurro, & Montaldi, 2020) suggests that detection of unexpected stimuli promotes pattern separation in the hippocampus, a computational mechanism that supports subsequent recollection (Azab, Stark, & Stark, 2014; Bakker, Kirwan, Miller, & Stark, 2008). Nevertheless, selective allocation of attention to the unexpected stimuli does not benefit familiarity discrimination of these stimuli. Instead the expected stimuli are the ones benefited by a boost in familiarity memory.

These findings may have important educational implications, in that learning objectives should optimally combine scenarios in which both expected and unexpected new bits of information are offered to learners. Therefore, a learning session may have optimal outcomes, in terms of memory retention, when information conforms to previously established rules. On the other hand, unexpected and surprising information, for example discussing a concept from a novel, unexpected or less obvious standpoint, can be beneficial for the creation of elaborative memories. These memories can be optimally benefitted from the creation of new associative links.

Taken together, the findings indicate that inconsistencies in the way expectation affects memory can be resolved by examining the type of memory that is formed and retrieved. Indeed, this tug of war between expected and unexpected stimuli is resolved by involving different encoding mechanisms. As discussed in the next section, this is closely demonstrated by differential pupil response patterns.

4.2. Type of novelty modulates pupil response patterns at encoding

The pupil response at encoding was closely related to subsequent memory type differentiating between missed, familiar and recollected stimuli. It also differentiated between multiple levels of memory strength, when the full scale of familiarity strength was used. However, the way the pupil response relates to subsequent memory was not unidirectional, but the pattern was modulated by expectation. Previous studies have provided inconsistent findings with respect to the direction of encoding-linked pupil responses, with some studies proposing increased dilation patterns as predicting successful memory formation (e.g., Papesh et al., 2012), while others finding that the degree of diminished dilation or constriction predicts successful memory formation (e.g., Kafkas & Montaldi, 2011; Naber et al., 2013).

The present findings unite these observations as they clearly suggest that the pattern of pupil response at encoding relates to the detection of different types of novelty. Therefore, unexpected stimuli are more likely to result in pupil dilation patterns and, importantly, the degree of this dilation predicts the strength and type of subsequent

memory. In contrast, expected stimuli result in greater pupil constriction patterns at encoding and, critically, the degree of this constriction relates to the strength and type of reported memory. An interesting observation is that differential memory advantage between expected and unexpected stimuli can occur in quick succession during memory formation within the same experimental session. Therefore, the encoding mechanisms that prioritise different types of novelty, rely on fast switching to accomplish flexible memory formation.

A free-viewing task was used at encoding as it is considered more appropriate for pupillometry studies exploring memory formation (see also extensive discussion in Kafkas & Montaldi, 2011). Considering the sensitivity of the pupil response to effortful processing and working memory load (Kahneman & Beatty, 1966), it is undoubtable that when a decisional task is employed, an effect on pupil response may be found, not necessarily because a memory for the stimulus is formed, but because a decision is being made. A recent study proposed that pupil response patterns at encoding (and especially pupil dilations) are not predictive of subsequent memory but only reflect decision urgency (Gross & Dobbins, 2020). The findings here show that this explanation is limited and certainly does not apply to all reported encoding-linked pupil effects. Effort associated with a cognitive task at encoding or the urgency of taking a decision, may infiltrate the pupil response, but they cannot be the sole explanation of the close link between pupil response and successful memory formation. Instead, the findings show that the most critical factor explaining previous inconsistencies in the way pupil response relates to subsequent memory is the type of novelty detected.

4.3. Implications for novelty mechanisms, brain networks and neurotransmission

What is the cognitive origin of these effects and what predictions can be made regarding their neural underpinnings? Expected stimuli are more likely to be encoded in a way that supports familiarity with the item, benefited by the detection of highly predictive outcomes. Consistent with the visual discrimination literature, where predictive cues result in better visual discrimination of stimuli (e.g., Bollinger, Rubens, Zanto, & Gazzaley, 2010; Posner, Snyder, & Davidson, 1980; Puri & Wojciulik, 2008; Trapp, Lepsien, Kotz, & Bar, 2016) expected items at encoding drive more fluent and efficient processing of information. Nevertheless, the degree of novelty, even in the case of expected outcomes, is sufficient to involve encoding mechanisms in the MTL (Kafkas & Montaldi, 2014). At the same time, this type of novelty detection presumably results in increased release of acetylcholine via the implication of the parasympathetic system (which is exclusively cholinergic) resulting in the observed pupil constriction patterns (Loewenfeld, 1999; Steinhauer et al., 2004) (for the role of acetylcholine in learning, memory and plasticity see Haam & Yakel, 2017; Hasselmo, 2006). Therefore, it is proposed here that the detection of expected novel stimuli enables encoding via cholinergic input to the MTL (and especially the cortical regions along the parahippocampal gyrus; see Kafkas et al., 2017) and makes them more easily recognisable on the basis of familiarity when encountered again.

On the other hand, unexpected stimuli are more surprising within a given context leading to aroused attention and enhanced exploratory behaviour (Barto et al., 2013). For example, unexpected stimuli are more likely to result in increased visual fixations (Kafkas & Montaldi, 2015a), and, like in the presented study, increased levels of pupil dilation (Kafkas & Montaldi, 2015a). At the neural level, contextual/unexpected novelty produces enhanced P300 event-related brain potential (ERP; Polich, 2007; Ranganath & Rainer, 2003), which also relates to pupil dilation (Murphy, Robertson, Balsters, & O'Connell, 2011) and has been linked to hippocampal function (Fonken, Kam, & Knight, 2020). Importantly, it has been shown to engage a functional loop involving the dopaminergic midbrain (especially the SN/VTA) and the hippocampus (Kafkas & Montaldi, 2015a; Lisman & Grace, 2005; Shohamy & Wagner, 2008; Wittmann, Bunzeck, Dolan, & Düzel, 2007). The adjacent locus coeruleus (LC), which is associated with the degree of phasic pupil dilation (Eldar, Cohen, &

Niv, 2013; Mather, Clewett, Sakaki, & Harley, 2016; Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014) has been linked to the detection of saliency, novelty and the attentional prioritisation of significance in the environment (Aston-Jones & Cohen, 2005; Markovic, Anderson, & Todd, 2014; Mather et al., 2016; Vazey, Moorman, & Aston-Jones, 2018). It is also a hotspot of norepinephrine (Szabadi, 2013; Walling, Brown, Milway, Earle, & Harley, 2011), as well as dopaminergic (Duszkiewicz, McNamara, Takeuchi, & Genzel, 2019; Kempadoo, Mosharov, Choi, Sulzer, & Kandel, 2016; Takeuchi et al., 2016; Wagatsuma et al., 2018) release to the hippocampus and the rest of the MTL. Therefore, it is proposed here that the pupil dilation as predictor of memory formation is linked to the detection of unexpected novelty and denotes dopaminergic and/or noradrenergic modulation (from SN/MTA and LC) of the hippocampus and the MTL cortex. This mechanism makes it more likely that unexpected stimuli are initially pattern separated and subsequently recollected.

Although in the present study expectation is proposed as a central factor defining the encoding mechanisms engaged when novel (unstudied) stimuli are presented, the proposal does not preclude effects of expectancy on processing non-novel stimuli (e.g., recently pre-familiarised items; for relevant evidence see Frank, Kafkas, & Montaldi, 2020; Kafkas & Montaldi, 2015a, 2018a). Therefore, the present findings and previous observations suggest that expectancy has a broad role in guiding memory encoding, reencoding as well as memory search and retrieval.

4.4. Conclusions

Taken together, the present findings demonstrate that the close link between pupil response and memory formation is sensitive to the type of novelty as defined by expectation. It is proposed that the relationship between either pupil dilation or constriction patterns, on the one hand, and the type of subsequent memory on the other, denotes the employment of different encoding mechanisms. These mechanisms are proposed to exert differential neurotransmitter control on memory formation regions of the brain and occur in quick succession in the service of flexible memory formation.

References

- Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience*, 28(1), 403–450.
<https://doi.org/10.1146/annurev.neuro.28.061604.135709>
- Azab, M., Stark, S. M., & Stark, C. E. L. (2014). Contributions of human hippocampal subfields to spatial and temporal pattern separation. *Hippocampus*, 24(3), 293–302.
<https://doi.org/10.1002/hipo.22223>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. L. (2008). Pattern Separation in the Human Hippocampal CA3 and Dentate Gyrus. *Science*, 319(5870), 1640–1642.
<https://doi.org/10.1126/science.1152882>
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge University Press.
- Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or Surprise? *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00907>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Bein, O., Livneh, N., Reggev, N., Gilead, M., Goshen-Gottstein, Y., & Maril, A. (2015). Delineating the effect of semantic congruency on episodic memory: the role of integration and relatedness. *PloS One*, 10(2), e0115624.
<https://doi.org/10.1371/journal.pone.0115624>
- Biel, D., & Bunzeck, N. (2019). Novelty Before or After Word Learning Does Not Affect Subsequent Memory Performance. *Frontiers in Psychology*, 10.
<https://doi.org/10.3389/fpsyg.2019.01379>
- Bollinger, J., Rubens, M. T., Zanto, T. P., & Gazzaley, A. (2010). Expectation-Driven Changes in Cortical Functional Connectivity Influence Working Memory and Long-Term Memory Performance. *Journal of Neuroscience*, 30(43), 14399–14410.
<https://doi.org/10.1523/JNEUROSCI.1547-10.2010>
- Bunzeck, N., & Düzel, E. (2006). Absolute coding of stimulus novelty in the human substantia nigra/MTA. *Neuron*, 51(3), 369–379. Retrieved from <http://www.sciencedirect.com/science/article/B6WSS-4KJD5DW-D/2/7b0b5c4dd4584088ab8fe4b58086e2d1>
- Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104(3), 268–294.
<https://doi.org/10.1037/0096-3445.104.3.268>
- Duszkiewicz, A. J., McNamara, C. G., Takeuchi, T., & Genzel, L. (2019). Novelty and Dopaminergic Modulation of Memory Persistence: A Tale of Two Systems. *Trends in Neurosciences*, 42(2), 102–114. <https://doi.org/10.1016/j.tins.2018.10.002>
- Eldar, E., Cohen, J. D., & Niv, Y. (2013). The effects of neural gain on attention and learning. *Nature Neuroscience*, 16(8), 1146–1153. <https://doi.org/10.1038/nn.3428>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Fenker, D. B., Frey, J. U., Schuetze, H., Heipertz, D., Heinze, H. J., & Düzel, E. (2008). Novel scenes improve recollection and recall of words. *Journal of Cognitive Neuroscience*, 20(7), 1250–1265.
- Fonken, Y. M., Kam, J. W. Y., & Knight, R. T. (2020). A differential role for human hippocampus in novelty and contextual processing: Implications for P300. *Psychophysiology*, 57(7). <https://doi.org/10.1111/psyp.13400>

- Frank, D., & Kafkas, A. (2020). Expectation-driven novelty effects in episodic memory. *Manuscript Submitted for Publication*.
- Frank, D., Kafkas, A., & Montaldi, D. (2020). Expectation violation dynamically engages a bottom-up encoding state. *BioRxiv*.
<https://doi.org/https://doi.org/10.1101/2020.12.15.422817>
- Frank, D., Montemurro, M. A., & Montaldi, D. (2020). Pattern Separation Underpins Expectation-Modulated Memory. *The Journal of Neuroscience*, 40(17), 3455–3464.
<https://doi.org/10.1523/JNEUROSCI.2047-19.2020>
- Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, 6(4), 187–214.
<https://doi.org/10.1080/17588928.2015.1020053>
- Gallucci, M. (2019). *GAMLj: General analyses for linear models*. Retrieved from <https://gamlj.github.io/>
- Geraci, L., & Manzano, I. (2010). Distinctive items are salient during encoding: Delayed judgements of learning predict the isolation effect. *The Quarterly Journal of Experimental Psychology*, 63(1), 50–64. <https://doi.org/10.1080/17470210902790161>
- Granholm, E., Asarnow, R. F., Sarkin, A. J., & Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology*, 33(4), 457–461. Retrieved from <http://www.sciencedirect.com/science/article/B6T1G-3P4TDWM-1B5/2/d8d8d5c518f3f80eee21b60068cbf8f2>
- Greve, A., Cooper, E., Tibon, R., & Henson, R. N. (2019). Knowledge is power: Prior knowledge aids memory for both congruent and incongruent events, but in different ways. *Journal of Experimental Psychology: General*, 148(2), 325–341.
<https://doi.org/10.1037/xge0000498>
- Gross, M. P., & Dobbins, I. G. (2020). Pupil dilation during memory encoding reflects time pressure rather than depth of processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000818>
- Haam, J., & Yakel, J. L. (2017). Cholinergic modulation of the hippocampal region and memory function. *Journal of Neurochemistry*, 142, 111–121.
<https://doi.org/10.1111/jnc.14052>
- Hasselmo, M. E. (2006). The role of acetylcholine in learning and memory. *Current Opinion in Neurobiology*, 16(6), 710–715. <https://doi.org/10.1016/j.conb.2006.09.002>
- Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity During Simple Problem-Solving. *Science*, 143(361), 1190–.
- Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth’s R² GLMM to random slopes models. *Methods in Ecology and Evolution*, 5(9), 944–946.
<https://doi.org/10.1111/2041-210X.12225>
- Joshi, S., & Gold, J. I. (2020). Pupil Size as a Window on Neural Substrates of Cognition. *Trends in Cognitive Sciences*, 24(6), 466–480.
<https://doi.org/10.1016/j.tics.2020.03.005>
- Kafkas, A., Mayes, A. R., & Montaldi, D. (2020). Thalamic-Medial Temporal Lobe Connectivity Underpins Familiarity Memory. *Cerebral Cortex*, 30(6), 3827–3837.
<https://doi.org/10.1093/cercor/bhz345>
- Kafkas, A., & Migo, E. M. (2009). Familiarity and recollection in the medial temporal lobe. *Journal of Neuroscience*, 29(8), 2309–2311. <https://doi.org/10.1523/jneurosci.5874-08.2009>
- Kafkas, A., Migo, E. M., Morris, R. G., Kopelman, M. D., Montaldi, D., & Mayes, A. R. (2017). Material Specificity Drives Medial Temporal Lobe Familiarity But Not Hippocampal Recollection. *Hippocampus*, 27(2), 194–209. <https://doi.org/10.1002/hipo.22683>
- Kafkas, A., & Montaldi, D. (2011). Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *The 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. (2014). Two separate, but interacting, neural systems for familiarity and novelty detection: A dual-route mechanism. *Hippocampus*, 24(5), 516–527. <https://doi.org/10.1002/hipo.22241>
- Kafkas, A., & Montaldi, D. (2015a). Striatal and midbrain connectivity with the hippocampus selectively boosts memory for contextual novelty. *Hippocampus*, 25(11), 1262–1273.
<https://doi.org/10.1002/hipo.22434>
- Kafkas, A., & Montaldi, D. (2015b). The pupillary response discriminates between subjective and objective familiarity and novelty. *Psychophysiology*, 52(10), 1305–1316.
<https://doi.org/10.1111/psyp.12471>
- Kafkas, A., & Montaldi, D. (2018a). Expectation affects learning and modulates memory experience at retrieval. *Cognition*, 180, 123–134.
<https://doi.org/10.1016/j.cognition.2018.07.010>
- Kafkas, A., & Montaldi, D. (2018b). How do memory systems detect and respond to novelty? *Neuroscience Letters*, 680, 60–68.
<https://doi.org/10.1016/j.neulet.2018.01.053>
- Kahneman, D. (1973). *Attention and effort* (Prentice-H). NJ.
- Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science*, 154(3756), 1583–1585. <https://doi.org/10.1126/science.154.3756.1583>
- Kempadoo, K. A., Mosharov, E. V., Choi, S. J., Sulzer, D., & Kandel, E. R. (2016). Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. *Proceedings of the National Academy of Sciences*, 113(51), 14835–14840. <https://doi.org/10.1073/pnas.1616515114>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13).
<https://doi.org/10.18637/jss.v082.i13>
- Lisman, J. E., & Grace, A. A. (2005). The hippocampal-VTA loop: controlling the entry of information into long-term memory. *Neuron*, 46(5), 703–713.
<https://doi.org/10.1016/j.neuron.2005.05.002>
- Loewenfeld, I. E. (1999). *The pupil: Anatomy, physiology and clinical applications* (Vol. 1). Boston, Oxford, Auckland, Johannesburg, Melbourne, New Delhi: Butterworth, Heinemann.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87(3), 252–271.
- Markovic, J., Anderson, A. K., & Todd, R. M. (2014). Tuning to the significant: Neural and genetic processes underlying affective enhancement of visual perception and memory. *Behavioural Brain Research*, 259, 229–241. <https://doi.org/10.1016/j.bbr.2013.11.018>
- Mather, M., Clewett, D., Sakaki, M., & Harley, C. W. (2016). Norepinephrine ignites local hotspots of neuronal excitation: How arousal amplifies selectivity in perception and memory. *Behavioral and Brain Sciences*, 39, e200.
<https://doi.org/10.1017/S0140525X15000667>
- Mayes, A. R., Montaldi, D., Roper, A., Migo, E. M., Gholipour, T., & Kafkas, A. (2019). Amount, not strength of recollection, drives hippocampal activity: A problem for apparent word familiarity-related hippocampal activation. *Hippocampus*, 29(1), 46–59.
<https://doi.org/10.1002/hipo.23031>
- Miller, A. L., & Unsworth, N. (2019). Variation in attention at encoding: Insights from pupillometry and eye gaze fixations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. <https://doi.org/10.1037/xlm0000797>
- Montaldi, D., & Mayes, A. R. (2010). The role of recollection and familiarity in the functional differentiation of the medial temporal lobes. *Hippocampus*, 20(11), 1291–1314.

- <https://doi.org/10.1002/hipo.20853>
- Murphy, P. R., O'Connell, R. G., O'Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain Mapping*, 35(8), 4140–4154. <https://doi.org/10.1002/hbm.22466>
- Murphy, P. R., Robertson, I. H., Balsters, J. H., & O'Connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology*, 48(11), 1532–1543. <https://doi.org/10.1111/j.1469-8986.2011.01226.x>
- Naber, M., Frässle, S., Rutishauser, U., Einhäuser, W., Frässle, S., Rutishauser, U., & Einhäuser, W. (2013). Pupil size signals novelty and predicts later retrieval success for declarative memories of natural scenes. *Journal of Vision*, 13(2), 11. <https://doi.org/10.1167/13.2.11>
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, 4(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Okuda, K., Højgaard, K., Privitera, L., Bayraktar, G., & Takeuchi, T. (2020). Initial memory consolidation and the synaptic tagging and capture hypothesis. *European Journal of Neuroscience*, ejn.14902. <https://doi.org/10.1111/ejn.14902>
- 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>
- Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, 6(2), 93–102.
- 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>
- Peinkhofer, C., Knudsen, G. M., Moretti, R., & Kondziella, D. (2019). Cortical modulation of pupillary function: systematic review. *PeerJ*, 7, e6882. <https://doi.org/10.7717/peerj.6882>
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109(2), 160–174. <https://doi.org/10.1037/0096-3445.109.2.160>
- Puri, A. M., & Wojciulik, E. (2008). Expectation both helps and hinders object perception. *Vision Research*, 48(4), 589–597. <https://doi.org/10.1016/j.visres.2007.11.017>
- Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, 4(3), 193–202. Retrieved from <http://dx.doi.org/10.1038/nrn1052>
- Rangel-Gomez, M., & Meeter, M. (2013). Electrophysiological analysis of the role of novelty in the von Restorff effect. *Brain and Behavior*, 3(2), 159–170. <https://doi.org/10.1002/brb3.112>
- Reichardt, R., Polner, B., & Simor, P. (2020). Novelty Manipulations, Memory Performance, and Predictive Coding: the Role of Unexpectedness. *Frontiers in Human Neuroscience*, 14. <https://doi.org/10.3389/fnhum.2020.00152>
- Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components. *Biometrics Bulletin*, 2(6), 110. <https://doi.org/10.2307/3002019>
- Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: overconfident estimates in mixed models. *Behavioral Ecology*, 20(2), 416–420. <https://doi.org/10.1093/beheco/arn145>
- Schmidt, S. R., & Schmidt, C. R. (2017). Revisiting von Restorff's early isolation effect. *Memory & Cognition*, 45(2), 194–207. <https://doi.org/10.3758/s13421-016-0651-6>
- Schomaker, J., & Meeter, M. (2015). Short- and long-lasting consequences of novelty,

- deviance and surprise on brain and cognition. *Neuroscience & Biobehavioral Reviews*, 55, 268–279. <https://doi.org/10.1016/j.neubiorev.2015.05.002>
- Schomaker, J., van Bronkhorst, M. L. V., & Meeter, M. (2014). Exploring a novel environment improves motivation and promotes recall of words. *Frontiers in Psychology*, 5, 918. <https://doi.org/10.3389/fpsyg.2014.00918>
- Shohamy, D., & Wagner, A. D. (2008). Integrating Memories in the Human Brain: Hippocampal-Midbrain Encoding of Overlapping Events. *Neuron*, 60(2), 378–389. Retrieved from <http://www.sciencedirect.com/science/article/B6WSS-4TPWYS3-N/2/babb4a89420131fa87ee67bbee5ca17f>
- Staresina, B. P., Gray, J. C., & Davachi, L. (2008). Event Congruency Enhances Episodic Memory Encoding through Semantic Elaboration and Relational Binding. *Cereb. Cortex*, 19(5), 1198–1207. <https://doi.org/10.1093/cercor/bhn165>
- Steinhauer, S. R., & Hakerem, G. (1992). The pupillary response in cognitive psychophysiology and schizophrenia. In D. Friedman & G. Bruder (Eds.), *Psychophysiology and Experimental Psychopathology: A Tribute to Samuel Sutton* (Vol. 658, pp. 182–204). New York: Academy of Sciences.
- Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology*, 52(1), 77–86. Retrieved from <http://www.sciencedirect.com/science/article/B6T3M-4BM92MF-C/2/978c1219756e3d43fae63adde0ad4365>
- Szabadi, E. (2013). Functional neuroanatomy of the central noradrenergic system. *Journal of Psychopharmacology*, 27(8), 659–693. <https://doi.org/10.1177/0269881113490326>
- Takeuchi, T., Duzskiewicz, A. J., Sonneborn, A., Spooner, P. A., Yamasaki, M., Watanabe, M., ... Morris, R. G. M. (2016). Locus coeruleus and dopaminergic consolidation of everyday memory. *Nature*, 537(7620), 357–362. <https://doi.org/10.1038/nature19325>
- The jamovi project*. (2020). Retrieved from <https://www.jamovi.org>
- Trapp, S., Lepsien, J., Kotz, S. A., & Bar, M. (2016). Prior probability modulates anticipatory activity in category-specific areas. *Cognitive, Affective, & Behavioral Neuroscience*, 16(1), 135–144. <https://doi.org/10.3758/s13415-015-0373-4>
- Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review*, 2(3), 387–390.
- Vazey, E. M., Moorman, D. E., & Aston-Jones, G. (2018). Phasic locus coeruleus activity regulates cortical encoding of salience information. *Proceedings of the National Academy of Sciences*, 115(40), E9439–E9448. <https://doi.org/10.1073/pnas.1803716115>
- 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, 130–140. <https://doi.org/10.1111/j.1469-8986.2007.00606.x>
- Wagatsuma, A., Okuyama, T., Sun, C., Smith, L. M., Abe, K., & Tonegawa, S. (2018). Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context. *Proceedings of the National Academy of Sciences*, 115(2), E310–E316. <https://doi.org/10.1073/pnas.1714082115>
- Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the von Restorff phenomenon. *Psychological Bulletin*, 63, 410–424.
- Walling, S. G., Brown, R. A. M., Milway, J. S., Earle, A. G., & Harley, C. W. (2011). Selective tuning of hippocampal oscillations by phasic locus coeruleus activation in awake male rats. *Hippocampus*, 21(11), 1250–1262. <https://doi.org/10.1002/hipo.20816>
- Wetzel, N., Einhäuser, W., & Widmann, A. (2020). Picture-evoked changes in pupil size predict learning success in children. *Journal of Experimental Child Psychology*, 192, 104787. <https://doi.org/10.1016/j.jecp.2019.104787>
- Widmann, A., Schröger, E., & Wetzel, N. (2018). Emotion lies in the eye of the listener:

- Emotional arousal to novel sounds is reflected in the sympathetic contribution to the pupil dilation response and the P3. *Biological Psychology*, 133, 10–17.
<https://doi.org/10.1016/j.biopsycho.2018.01.010>
- Wittmann, B. C., Bunzeck, N., Dolan, R. J., & Duzel, E. (2007). Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *NeuroImage*, 38(1), 194–202. Retrieved from <http://www.sciencedirect.com/science/article/B6WNP-4P8B0SM-3/2/45d5a019f2437e771651b2a5cda419a8>
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language*, 46(3), 441–517.
<https://doi.org/10.1006/jmla.2002.2864>
- Yonelinas, A. P., & Jacoby, L. L. (1995). The Relation between remembering and knowing as the bases for recognition - Effects of size congruency. *Journal of Memory and Language*, 34(5), 622–643.