

A predictive account of how novelty influences declarative memory

Jörn Alexander Quent^{a,*}, Richard N. Henson^{a,b}, Andrea Greve^a

^a MRC Cognition and Brain Sciences Unit, University of Cambridge, United Kingdom

^b Department of Psychiatry, University of Cambridge, United Kingdom

ARTICLE INFO

Keywords:

Novelty
Memory
Predictive coding
Prediction error
Surprise
PIMMS

ABSTRACT

A rich body of studies in the human and non-human literature has examined the question how novelty influences memory. For a variety of different stimuli, ranging from simple objects and words to vastly complex scenarios, the literature reports that novelty improves memory in some cases, but impairs memory in other cases. In recent attempts to reconcile these conflicting findings, novelty has been divided into different subtypes, such as relative versus absolute novelty, or stimulus versus contextual novelty. Nevertheless, a single overarching theory of novelty and memory has been difficult to attain, probably due to the complexities in the interactions among stimuli, environmental factors (e.g., spatial and temporal context) and level of prior knowledge (but see Dusz-kiewicz et al., 2019; Kafkas & Montaldi, 2018b; Schomaker & Meeter, 2015). Here we describe how a predictive coding framework might be able to shed new light on different types of novelty and how they affect declarative memory in humans. More precisely, we consider how prior expectations modulate the influence of novelty on encoding episodes into memory, e.g., in terms of surprise, and how novelty/surprise affect memory for surrounding information. By reviewing a range of behavioural findings and their possible underlying neurobiological mechanisms, we highlight where a predictive coding framework succeeds and where it appears to struggle.

1. Introduction

Events that are new, different or unusual often “stick in your mind”. For example, imagine you live and work in an urban area and you take the same route to work in your car every morning. One day you see a flock of sheep blocking the road on your commute. You will probably remember this event for a long time, while the other countless times you have driven down that road are forgotten. Numerous studies have confirmed this observation, namely that if we experience a novel event in a familiar context, we tend to store and remember this event more easily. However, if you happen to live in the countryside close to a sheep farm, your experience might be quite different: because you see flocks of sheep quite frequently, you might experience just another regular commute to work that is hardly memorable. Thus, the experience of novelty is not ‘absolute’ and cannot be defined independent of the observer; rather it is driven by what an individual expects to experience compared to what they actually encounter. This comparison between expected and experienced outcomes parallels the computation of a prediction error (PE) in many theories of learning. According to such theories, we continuously generate expectations about our environment,

and update those predictions when they are wrong, i.e., when a PE occurs. While this role of PE in learning is well established in experiments on non-declarative memory (e.g., in conditioning and associative learning in animals), its role in human declarative memory (i.e., conscious, verbalisable memory; Squire, 2004) is less well established. In line with a recent review by Reichardt et al. (2020), we propose that predictive coding theories in general, and the ‘Predictive Interactive Multiple Memory Signals’ (PIMMS) proposal of Henson and Gagnepain (2010) in particular, provide a powerful framework to help us understand how novelty influences declarative memory. Going beyond Reichardt et al. (2020), we show how PIMMS reveals multiple different types of novelty, including situations that lead to poor memory, and how it allows for PE at different levels of representation, which might map onto different types of memory (e.g., recollection versus familiarity) that involve distinct brain regions. In doing so, we also reveal empirical findings that are not addressed by PIMMS, such as the effect of novelty on surrounding information, and a memory advantage for expected information (with little apparent PE).

* Corresponding author at: MRC Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge CB2 7EF, United Kingdom.
E-mail address: alex.quent@mrc-cbu.cam.ac.uk (J.A. Quent).

<https://doi.org/10.1016/j.nlm.2021.107382>

Received 15 July 2020; Received in revised form 8 November 2020; Accepted 10 January 2021

Available online 18 January 2021

1074-7427/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

2. Novelty and prediction errors

The question of how we remember novel information has inspired numerous studies over the past decades. Berlyne (1960) was the first to introduce the term *absolute novelty* for something that has never been encountered before, as distinct from *relative novelty*, which describes novel arrangements of familiar elements. More recent accounts use the term *stimulus novelty* to refer to absolute novelty, and contrast this with *contextual novelty* (Ranganath & Rainer, 2003; Schomaker & Meeter, 2015), which arises when a familiar item is encountered in an unexpected context (more akin to Berlyne's relative novelty). Regardless of the terminology, what these concepts have in common is that both arise against a backdrop of (the presence or absence of) prior expectations. The PIMMS framework offers one way to think about the relationship between the precision of predictions and the uncertainty of sensory input.

According to PIMMS, memory and perception arise within multiple levels of a processing hierarchy, where higher levels are constantly predicting the activity in lower levels, and the difference between the predicted and actual activity (at a single moment in time) – the PE – is fed back from lower to higher levels. For a given layer, the predictions from the level above are equivalent to a “prior” probability distribution (in a Bayesian sense), whereas the activity profile (produced from the level below) is equivalent to the evidence or “likelihood”, while the PE is the *divergence* between these two distributions (the summed area of no overlap)¹. The size of the PE then determines how well a new event is encoded into memory.

Let us consider a level representing the current items perceived (e.g., objects), which may or may not be predicted from a higher level representing the context (e.g., environment). Take the example of sheep encountered in an urban environment: the context is predicting houses, road-signs, other cars, etc., whereas the sensory input is indicating sheep. This corresponds to a prior and likelihood that are quite precise but differ in their modes (Fig. 1A). This results in high PE, which causes strong encoding of the event. We refer to such situations as “surprise”, in keeping with other related work (Reichardt et al., 2020). In fact, this surprise could occur at multiple levels of the processing hierarchy. When it occurs between contextual predictions about familiar items, we will call it “context surprise”. However, our knowledge of familiar items (e.g., sheep) allows us to make predictions about the perceptual features (e.g. four legs, white wool, etc.) that comprise those items. When one or more features differ from those expected (e.g., a pink sheep), then PE occurs at this lower level of the hierarchy. We call this “item surprise” (Fig. 1B); similar to Berlyne's concept of relative novelty.²

Now consider another situation, where you enter a new environment (a context that you have not experienced before), then you have minimal predictions about what items to encounter, such that the prior is imprecise, or in the extreme case, “flat”. We will call this “context novelty”. When you do encounter a familiar object (such that the likelihood is precise), there is still a PE (though not as great as in Fig. 1A/B), which can lead to reasonable memory encoding (Fig. 1C). For example, imagine you had always lived on a remote sheep farm, and entered a city

for the first time, where you do not quite know what to expect. If you encountered a flock of sheep, this might be somewhat surprising, even if not as surprising as for the urban commuter considered above.

Or consider a situation where you encounter an item you have never seen before. In this case, the likelihood is flat instead – i.e., you do not know how to interpret the sensory evidence. We call this “item novelty”. When you encounter such an unknown item in a familiar environment, where you are expecting other objects instead (i.e. a precise prior, Fig. 1D), then there is again a small PE. This would correspond to the urban commuter encountering the sheep, but in this case, having no prior knowledge of animals like sheep at all.

Finally, in the case when the prior and the likelihood overlap closely (Fig. 1E), so that PE is low, there is no need to do any learning (i.e., no need to waste resources re-encoding what is already known). However, a low PE can also emerge from a flat prior and flat likelihood, as in Fig. 1F. Importantly, this situation of “complete (or maximal) novelty” (e.g., encountering unknown objects in an unknown environment) is actually predicted to produce negligible, rather than “maximal”, learning.

PIMMS originally distinguished between at least three levels that differ in their representational content: the “episodic” system, at the top of the hierarchy, which stores associations between spatiotemporal contexts and items (objects); a “semantic” level, which contains knowledge about individual items and their corresponding configuration of perceptual features; while lower in the hierarchy, “perceptual” systems represent individual, modality-specific features of the stimuli.

The key function of the episodic system is to optimize the predictability of an item occurring in a particular spatiotemporal context, i.e., store context-item associations. When we encounter a familiar item in a context that is different from that expected, the ensuing PE induces learning of more accurate associations between episodic and semantic representations. This is the type of learning that enables memory of the spatiotemporal context in which an item occurred, or what has been called “recollection” (Mandler, 1980; Montaldi & Mayes, 2010; Yonelinas, 2002). This episodic level is associated with the hippocampus and other brain structures linked with episodic memory (Aggleton & Brown, 1999; Mayes et al., 2007; Moscovitch, 1995; Moscovitch et al., 2016; Scoville & Milner, 1957; Squire, 1992). Note that PEs can, and normally do, arise automatically, based on prior knowledge triggered by perceptual inputs, i.e., predictions are rarely intentional or effortful. However, whereas Morris (2006) claimed that attended experiences are automatically recorded by the hippocampus, but will normally fade and be lost, we claim that PE modulates the degree of encoding and therefore determines whether an experience will be available as a lasting memory.

The semantic level on the other hand stores information about familiar items, and predicts which features are expected on the basis of a given item being present. Note that there is a bidirectional flow of information between all systems, so not only do currently active item representations make predictions about associated perceptual features, but currently active features also influence which item representations remain active (i.e., perception is a dynamic competition across all systems in order to minimise overall PE). The semantic system is associated with anterior temporal lobe regions, including perirhinal cortex, and it is the strengthening of item-feature associations that enables the feeling of “familiarity” (rather than recollection), which sometimes accompanies recognition memory (Mandler, 1980; Montaldi & Mayes, 2010; Yonelinas, 2002).

There are probably numerous levels of intermediate perceptual representations, depending on the modality, but for simplicity we consider posterior temporo-occipital regions associated with perception of visual stimuli. However, since we focus here on declarative memory, we will concentrate on the episodic and semantic systems.

Given this brief overview of PIMMS, we summarise different types of novelty/surprise in Table 1. With these definitions in mind, we will first see how well they can capture the human behavioural literature. Then we will consider how they might relate to neurobiological mechanisms, such as those studied in the animal literature.

¹ More precisely, the PE that drives learning is the divergence between the prior and the posterior, on the assumption that the initial PE between the prior and the likelihood is what drives perception (activity changes), which seeks to minimise this PE over a few hundred milliseconds (producing the posterior distribution), leaving the residual PE that drives synaptic change instead (Henson & Gagnepain, 2010). For simplicity though, we assume here that the posterior is close to the likelihood, such that the qualitative implications are the same.

² Note that that predictions in PIMMS are generally implicit (automatic, unconscious), though they could also be explicit (conscious) and/or generated intentionally. PIMMS does not currently distinguish between these, and makes the same (experimental) prediction of improved memory when PE is high, regardless of whether the (theoretical) prediction is implicit or explicit.

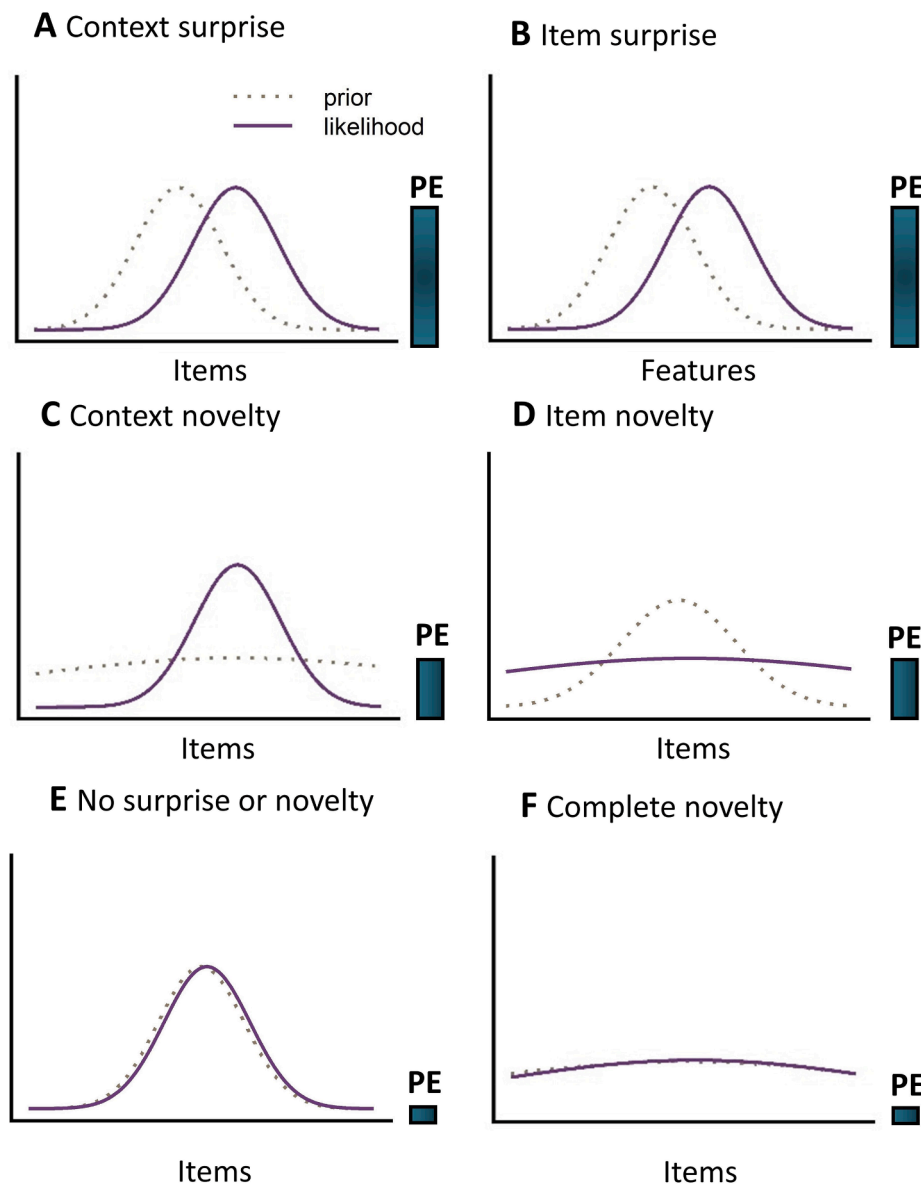


Fig. 1. Schematic illustration of prior and likelihood distributions in the “predictive interactive multiple memory signals” framework (PIMMS). Distributions reflect activity in a layer of topographically organised neurons, where the x-axis captures similarity between items in the semantic level or similarity between features in the perceptual layer (depending on the panel). The dotted line represents the prior predictions from the “higher” level, whereas the solid line represents the likelihood distribution, input from the level below (ultimately the sensory input). The PE, which drives learning, is the divergence between these two probability distributions, whose magnitude is illustrated at the bottom right of each panel. Panel A: precise prior and precise likelihood for items in a certain context, but with different modes (context surprise). Panel B: the same as Panel A, except the precise prior and likelihood refer to features of an object in lower perceptual levels (item surprise). Panel C: Flat prior and precise likelihood (context novelty). Panel D: precise prior and flat likelihood (item novelty). Panel E: precise prior and precise likelihood with the same mean (leading to no PE or learning; no surprise or novelty). Panel F: flat prior and flat likelihood, a combination of context novelty and item novelty (or “complete novelty”), but one predicted to show no PE or learning.

3. Context surprise

Context surprise is the most studied form of novelty in the human psychological literature. Going back to our initial example, the context we are in (being on the way to work) predicts the presence of certain objects and events. If we now encounter something that is not predicted, like the sheep, then the resulting difference between prediction (the prior) and observation (the likelihood), give rise to a PE, which triggers the formation of an episodic memory for this unexpected event.

3.1. Expectations arising from shared episodic context

We start with what has been called the classic ‘novelty effect’ in episodic memory (Tulving & Kroll, 1995; though first mentioned by Kinsbourne & George, 1974). In this paradigm, the participant is familiarised with a list of random words. In a second “critical” phase, they see a list of new words, intermixed with some of the familiarised words. Finally, they are presented with a third list of words, and asked to recognise any that came from the critical phase (Fig. 2A). The common finding is that the new words are better recognised than the familiar words, despite the fact that they were presented fewer times in total

(note, it is critical that participants are asked to recognise items specifically from the critical phase; if participants are instructed to recognise items studied in either list, memory is better for items that were repeatedly presented (Kafkas & Montaldi, 2015a; Kim et al., 2012).

However, the cause of this ‘classic’ novelty effect has been challenged by alternative explanations, other than a special role for novelty in encoding, such as distinctiveness or source confusion at retrieval (Åberg & Nilsson, 2001; Dobbins et al., 1998; Greene, 1999; Poppenk et al., 2010). In general, these authors attribute the difference in recognition to an impairment for familiar stimuli owing to contextual interference, e.g., “Did I see this stimulus in the familiarisation phase (first list) or study phase (second list)?”, rather than a benefit for the non-familiarised (novel) stimuli, which produces increased false alarm rates and hence overall decreased old/new discrimination. However, other studies controlled for confounds like interference, and reinforced the conclusion of Tulving and Kroll (1995) conclusion. Åberg and Nilsson (2003), for instance, reported a novelty effect for high confidence responses that showed both increases in hit rates and decreases in false alarm rates for novel items, which suggests the effect cannot be explained fully by reduced retrieval accuracy for familiar items, but rather enhanced encoding of novel items. Similarly, Kormi-Nouri et al.

Table 1
Glossary with the main forms of surprise and novelty.

| Term | Definition | PIMMS | Example |
|------------------|---|---|---|
| Context surprise | When familiar items occur in an unexpected context. Has previously been called “contextual novelty” (Ranganath & Rainer, 2003; Schomaker & Meeter, 2015); though used differently here (see Context Novelty below) | A strong prior from the context level to the item level is accompanied by strong but divergent evidence from item level (Fig. 1A) | An urban commuter encounters a flock of sheep in the city. |
| Item surprise | When a familiar object that has one or more unexpected features. Analogous to what Berlyne (1960) called relative novelty. | A strong prior from the item level to the feature level is accompanied by strong but divergent evidence for one or more features (Fig. 1B) | A sheep farmer encounters a pink sheep. |
| Context novelty | The context is so unfamiliar that you do not know what to expect | A flat prior from the context level to item level, i.e., few predictions about the kind of items present. Note that whether or not PE occurs depends on whether the likelihood is precise or also flat (cf. Fig. 1C vs 1F) | A sheep farmer who has never visited a city before. |
| Item novelty | Items that have not been encountered before. Analogous to what Berlyne (1960) called absolute novelty, or more recent definitions of “stimulus novelty” (Ranganath & Rainer, 2003; Schomaker & Meeter, 2015). | Flat evidence at the item level, i.e., the features present do not activate a unique item representation. Note that whether or not PE occurs depends on whether the prior is precise or also flat (cf. Fig. 1D vs 1F). | An urban commuter encounters a sheep having never seen one before. |
| Complete novelty | Unknown items encountered in an unknown context. This is a combination of context novelty and item novelty. | Flat prior and flat likelihood (Fig. 1F) | An urban commuter encounters a sheep on a farm, having never been to a farm or seen a sheep before. |

(2005) presented distinct encoding tasks to minimize source confusion at retrieval, and still observed better memory for novel over familiar words, consistent with the novelty-encoding hypothesis.

How is this novelty-encoding effect explained by PIMMS? One possibility relates to predictions made by the temporal context. When you encounter the list of random words in the familiarisation phase, you are unable to predict the next word. However, in the critical phase, you notice that some of the words are repeated from the familiarisation phase, so you might start to expect further repetitions. Indeed, if novel items become less expected, they will elicit a greater PE than the familiar words, and therefore become more strongly associated with the context of the critical study phase (hence an example of context surprise). Therefore, when finally asked to recognise words specifically from that phase, you are better able to do so. If this interpretation is correct, then

the advantage of these “novel” words should depend on the strength of expectation for repeated words, which could be tested by manipulating the ratio of familiarised to new words in the critical phase. This prediction is consistent with Kafkas and Montaldi (2015a), who showed that the novelty effect reverses when the previously presented words are rare, i.e., it is the unexpectedness rather than relative familiarity that determines memory.

3.2. Expectations arising from shared semantics

Another long-standing line of research also uses repetition to build up expectations within the course of the experiment. However, rather than using temporal context to drive the expectation of pre-familiarised items, as in Tulving and Kroll (1995)’s classic novelty effect above, these studies repeat the same class or type of items, thereby generating expectations about the category of the next item (even if not the specific exemplar). Early studies used lists of simple stimuli such as letters, numbers or words, and found memory was superior for deviant (or “oddball”) items than for other items (at the same position in other lists) that were congruent with the context generated by the list (Fig. 2B). This finding is also known as ‘von Restorff effect’, since in one of her seminal studies, von Restorff (1933) presented four pairs of syllables along with one pair of symbols, numbers, letters or patches of colour. The pairs that only occurred once were better recalled immediately afterwards. Although better recall performance is generally thought to be supported by recollection, consistent with PIMMS in terms of contextual predictions at the episodic level, other studies that used recognition performance instead reported stronger familiarity as well (e.g. Kishiyama & Yonelinas, 2003; Rangel-Gomez & Meeter, 2013; von Restorff, 1933). According to PIMMS, increased familiarity would only be expected if the deviant items were also surprising in their perceptual form (see Item Surprise below).

3.3. Expectations arising from trained associations or rules

Other studies have varied the strength of predictions by training associations or relations between items. For example, Greve et al. (2017) repeatedly paired a scene with a category of words (Experiment 1) or a specific face (Experiment 2) during a training phase, to establish predictions for what was associated with each scene. In a critical encoding phase, the scene was paired once with a new word/face, and then memory for that new associate was tested in a final cued recall phase (where the cue was the scene and three response options were given, all from the encoding phase, but only one of which was paired with the cue; see Fig. 2C). Memory for the new association was better when the prediction was violated (Experiment 1) or when the prediction that was violated was more precise (Experiment 2), directly supporting a prediction error account.

In a later study, Greve et al. (2019) trained participants with simple rules that related two sets of objects, and found that events (i.e., individual trials with a specific number of each type of object) which violated expectations were better recognised than events that had weaker expectations. This converges with finding from Kafkas and Montaldi (2018a), who asked participants to associate symbols with categories of objects (living/non-living) and reported increased levels of recollection for objects that were unexpected. In another study, using the mismatch paradigm (see section Neuroimaging studies of novelty and surprise in humans; Fig. 2G), it was reported that memory was better for objects when their occurrence violated the sequence shown immediately before (Chen et al., 2015).

While the above studies are consistent with a general pattern of PE driving declarative memory, they also illustrate a limitation of PIMMS, in that it remains unclear what is generating these predictions, since the predictions from trained rules or item-item associations do not obviously come from the spatiotemporal context, nor from (pre-existing) semantic knowledge. The fact that the greater PEs lead to better recollection (e.g.,

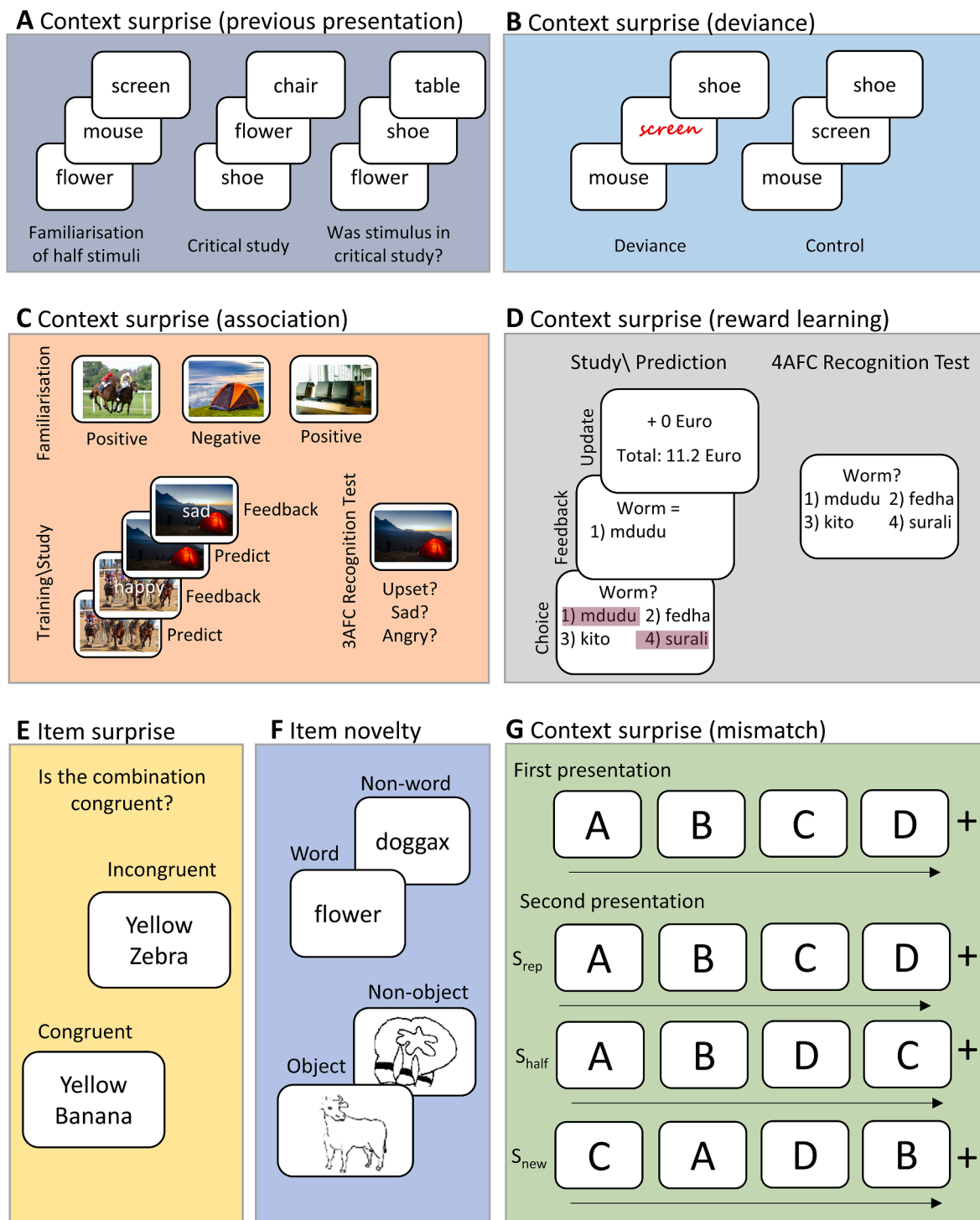


Fig. 2. Illustration of key paradigms. Panel A: Design by [Tulving and Kroll \(1995\)](#): pre-familiarised and novel items are presented intermixed at critical study for which recognition memory was later tested. Panel B: Design of von Restorff/distinctiveness paradigm: items of the same type/class are presented in lists together with a conceptually or perceptually deviant item (e.g. different font type/colour), memory for which can be compared to an item same position in a control list without deviants. Panel C: Rule based design by [Greve et al. \(2017\)](#): At study new scenes were paired with new words, which had the same valence as expected from a previous familiarisation phase (low PE) or the opposite valence (high PE). A forced-choice memory test matched target and foils to be of same valence and equally familiar. Panel D: Reward PE design by [De Loof et al. \(2018\)](#): one, two or four Swahili words are presented as options from which the rewarded word is selected, which manipulated the size of RPE. Panel E: Design by [Reggev et al. \(2018\)](#): Judging whether a noun-adjective is congruent and subsequently testing memory for the nouns. Panel F: Item novelty assessed by presenting objects vs. non-objects or words vs. non-words in [Kroll and Potter \(1984\)](#). Panel G: Mismatch design from [Kumaran and Maguire \(2006\)](#): sequences of objects were represented twice, wherein the second presentation, the order of objects was either unchanged (S_{rep}), changed after the first half (S_{half}) or completely new (S_{new}).

Kafkas & Montaldi, 2018a) suggests they involve the episodic system, but another possibility is that a large PE, regardless of where the prediction comes from, always results in better episodic encoding of the event associated with that error (see below). More generally, it seems likely that multiple predictions, e.g., about which objects will be present in an environment, can arise simultaneously from different sources: long-established semantic knowledge about certain environments (e.g., restaurants), recent rules or patterns observed (e.g., in “situation models” or working memory, like in the Greve et al. (2019), study) or even transient or established associations between objects that tend to co-occur. Distinguishing the roles of these different types of prediction, rather than attributing all to a hippocampal representation of “context”, may be an important future research direction, including the potential for different brain regions being involved.

3.4. Expectations arising in reinforcement learning

Another method is to establish predictions for a reward following a stimulus. In the human literature on such “reward PE” (RPE), a stimulus (or class of stimuli) is typically repeatedly paired with a reward (e.g., money), so that future presentations of that stimulus lead to the anticipation of reward, and a RPE is triggered if the reward is not delivered (Fig. 2D). Interestingly, thus far we have been talking about *unsigned* PE (in the context of PIMMS), but with RPE, because the reward has a direction (more or less), one can additionally distinguish *signed* RPE, i.e., positive when the actual reward is bigger than the predicted award, and negative when the actual reward is less. Although the literature on RPE has typically focused on non-declarative learning (e.g. O’Doherty et al., 2004), RPE can also affect declarative memory. De Loof et al. (2018), for example, asked participants to learn Dutch-Swahili word pairs, in which they varied the number of Swahili alternatives for each Dutch word (affecting the precision of the prediction) and the monetary reward (to produce either positive or negative RPE). The data revealed a positive, linear relationship between signed RPE (from negative to positive) and memory performance (see also Calderon et al., 2020). Jang et al. (2019) reported a similar result, though other studies have argued that memory performance is better predicted by unsigned RPE, i.e., its magnitude regardless of sign (Rouhani et al., 2018; Rouhani & Niv, 2019). The relationship between the sign of a RPE and subsequent memory therefore requires further investigation, which might include for example studies in which the reinforcement is aversive rather than rewarding (e.g., when a cue predicts a painful shock rather than money). It is also important to keep in mind the type of memory being tested: whether it is an association between a stimulus and a reward (or an association between two items), which might be declarative and even established on a single trial, but not necessarily episodic in the sense of being bound to a specific spatiotemporal context (e.g., trial in the experiment). Alternatively, it is possible, as suggested above, that a large PE triggers episodic encoding, in addition to strengthening associations between items and/or reinforcers.

3.5. Expectations arising from real-world knowledge

In the studies discussed so far, predictions were constructed ‘online’ within the course of the experiment and the results generally endorse the view that contextual surprise enhances memory. In real life, however, context surprise is more likely to arise from expectations that were established over an extended period of time, and some studies draw on this real world knowledge as a source of predictions.

For example, Pine et al. (2018) had participants learn factual information (e.g. “How many weeks did the Falklands war last?”) and found better memory for trials that violated participants’ prior expectations. Similarly, Foster and Keane (2019) presented stories that contained either well-known surprises (e.g. realising that a wallet is missing when walking down the street), less well-known surprises (e.g. realising that a belt is missing in the same scenario) or no surprises at all. The harder it

was to intuitively explain a surprise, the better the story was subsequently remembered, which fits well with a PE account.³ Similarly, Prull (2015) presented participants with everyday objects in typical or atypical locations, which were either consistent or inconsistent with their schema, i.e., prior knowledge about the structure of the world. The authors showed that recognition accuracy was better for schema-inconsistent trials, and this advantage was driven by recollection. Interestingly, recent studies have started to explore more naturalistic stimuli, such as video clips, whose continuous nature is an even better match of real life experience. In a study by Straube et al. (2014), for example, participants viewed video clips of an actor verbalising a sentence and concurrently performing a congruent gesture, an unrelated gesture or no gesture at all. Sentences with surprising (unrelated) gestures were better remembered than sentences with no gestures. More recent studies have used immersive virtual reality to demonstrate better memory for objects that occur in unexpected locations based on pre-experimental knowledge for where those objects should occur, e.g., within a virtual kitchen (Quent & Henson, 2018).

4. Item surprise

An example of item surprise would be encountering a pink sheep, i.e., a familiar object that has most of its expected features except for a few unexpected ones (Fig. 1B). According to PIMMS, this PE between the semantic and feature levels should improve encoding of this new object (in contrast to an unfamiliar animal, which happened to be pink). However, this is at odds with some studies in the literature that report better memory for item-feature binding that is congruent with pre-experimental knowledge. Staresina and Davachi (2006), for example, presented colour–nouns associations and when participants were instructed to create a mental image that combines both (e.g. yellow banana, pink elephant etc.), they were able to recall more plausible than implausible combinations at test. In a similar vein, Reggev et al. (2018) found superior memory for nouns that were previously paired with adjectives consistent with prior expectations (e.g. yellow banana), compared to those in pairs that violated such expectations (e.g. purple banana; Fig. 2E). Some of these cases where memory is better for items that are congruent with prior knowledge (rather than incongruent) might be explained by prior knowledge improving recall at test, e.g., to generate a typical banana from semantic memory, think of its colour as yellow, and then recognise that a yellow banana was presented (a so-called generate-and-recognise strategy; Watkins & Gardiner, 1979). This potential confound can be addressed by presenting response options that are equally congruent or equally incongruent with prior knowledge, e.g. two-alternative forced choice for a yellow banana vs. a green banana, or a purple banana vs. a red banana (see e.g. Quent & Henson, 2018).

These situations demonstrate that novelty/surprise is not the only factor that drives declarative memory encoding, and that sometimes, congruency with prior knowledge (or a “schema”) can boost memory encoding too, even though this is a situation with low PE (as in Fig. 1E). This cannot be explained by PIMMS, and has led some authors to propose that a separate brain system (involving the medial prefrontal cortex) supports the rapid learning of new information that is congruent with prior knowledge (van Kesteren et al., 2012). The existence of two types of learning (one based on large PE and one based on small PE) predicts that memory should be a U-shaped function of expectancy (or schema-congruency), where the highly expected and highly unexpected

³ Interestingly, this advantage for context surprise appears to be moderated by the degree of individual curiosity, according to Kang et al. (2009), who asked participants to firstly guess the correct answer to a question and secondly rate how curious they were to learn the true outcome. Recall accuracy for incorrectly guessed responses, i.e. novel or surprising facts, increased linearly with curiosity (also see Gruber & Ranganath, 2019).

extremes lead to better memory than the central point of minimal (un)expectedness (a prediction recently confirmed by Greve et al. (2019)). The important implication of this U-shaped function for present concerns is that, when just two conditions are compared (as in many of the studies above), whether memory is better for expected or unexpected information will depend on where those conditions fall on the U-shaped function: if they fall more towards the unexpected extreme, then the condition with greatest unexpectedness will produce better memory, whereas if they fall more towards the expected extreme, then the condition with greatest expectedness will produce better memory. Thus, designs with at least three levels of expectancy measures (i.e. unexpected, neutral and expected) are needed, to appropriately sample both extremes of the U-shape function.

5. Context novelty

Context novelty, rather than context surprise, refers to a novel context that makes no clear predictions. This describes the situation in many laboratory studies of memory for random lists of items, where there are minimal predictions about each item before it appears. Though PE is not as high as in the cases of surprise above, where there are (incorrect) predictions, PIMMS still predicts higher PE for familiar than novel items (since former have a more precise likelihood; Fig. 1C).

A different paradigm, but one whose results can be explained in a similar manner, is that used by Kaula and Henson (2020), where some items (faces) were pre-familiarised (primed), and then intermixed with new faces in a second critical phase, before memory for primed and new faces was compared in a final recognition phase (see also Experiment 3 of Greve et al., 2017). While this paradigm is similar to the Tulving and Kroll (1995) paradigm shown in Fig. 2A, an important difference is that Kaula and Henson (2020) paired each face with a unique context (scene) in the critical study phase, rather than the common “temporal” context assumed to apply in Tulving & Kroll’s paradigm. Since the scene-face pairing was arbitrary, the scene imposed no prior on the specific face, whereas it can be argued (as in Context Surprise section above), that the temporal context in the Tulving & Kroll paradigm did impose a prior towards expecting familiar items (depending on the ratio of familiar to new items). This may explain the opposite findings: whereas Tulving and Kroll found better memory for the new items (because they violated the expectation for familiar items), Kaula & Henson found better memory for the familiar items, in the form of better associative memory for the face-scene pairing. The latter is consistent with a greater PE for the primed than new items, owing to a more precise likelihood (in the presence of a flat prior; Fig. 1C). However, further work is needed to distinguish this PIMMS account from an attentional resource account, where priming results in less attention being devoted to processing the item, and more attention being devoted instead to processing its context (background scene; see Kaula & Henson, 2020, for further discussion).

6. Item novelty and complete novelty

The converse situation to context novelty is item novelty, where a completely new item is encountered that has no pre-existing semantic representation, i.e. a flat likelihood. If the context also makes no predictions, i.e., the prior is also flat (i.e., the case of “complete novelty” in Fig. 1F), then there should be minimal encoding. This is consistent with a number of empirical studies that use random lists of items (i.e., contexts with no strong priors) and compare memory for known vs. unknown items, such as familiar vs. unfamiliar symbols (Cycowicz, 2019; Cycowicz & Friedman, 2007), non-words vs. words (Belleville et al., 2011; Gardiner & Java, 1990; Perfect & Dasgupta, 1997) or unknown vs.

known melodies (Belleville et al., 2011; Java et al., 1995). As expected from PIMMS, episodic memory for the unknown items is worse.⁴ In the situation where the likelihood is flat but there is a precise prior, i.e., when unknown items are presented in a context that predicts certain known items, PIMMS predicts there should be some encoding driven by the PE (Fig. 1D). We are not aware, though, of any experiments that have compared this situation to the complete novelty situation of a flat prior.

Returning to the situation of complete novelty (Fig. 1F), this is interesting because PIMMS predicts little encoding, despite maximal novelty. According to PIMMS, the sensory systems (e.g. occipito-temporal cortex) are able to represent the perceptual details of completely novel items (to the extent that the incoming sensations can be processed at all), but the lack of prior knowledge in the semantic or episodic level prevents any effective predictions. While the amount of encoding in such situations may be minimal, it is clear that we must be able to learn something, in order to “bootstrap” learning of new items in new contexts. However, it may be that many such learning episodes are needed before predictions can be established, so that memory for any single episode is poor. Furthermore, other factors may also come into play in these situations of maximal novelty, such as increased attention/arousal, which may affect encoding of other information nearby in time, as we discuss next.

7. Effects of novelty on memory for surrounding information

So far, we have talked about memory for the novel information itself. In other situations, novelty/surprise may impact memory for additional types of incidental information that surround novel information (in time and/or space), even if that other information is relatively unrelated and not necessarily novel itself. Let us return to our example in which you encounter a flock of sheep on your way to work, which was highly surprising. There is good evidence that you will remember this event better than other morning commutes. However, what is it that will be stuck in your memory? Is it just the unusual event of seeing sheep, or will you show a benefit for other contextual information that was present at the same time, such as what song was playing in the radio, what clothes you were wearing or whether the sun was shining? In addition, are you also more likely to remember other moments that happened right before or after the encounter of the sheep, such as whether you did the school run that morning or with whom you had a meeting once you arrived at work?

7.1. Effects at the level of source information

One related question for PIMMS is whether PE only enhances learning of associations between the predictor (e.g., item) and the predicted (e.g. features) – i.e. learning that is “local” to the cause of the PE – or does it trigger a “global” increase in learning which also enhances memory for other incidental information that happens to co-occur with a PE? An example of memories that contain rich incidental, episodic information are “flashbulb” memories, which lead to clear recollection of the context surrounding a particular event, such as where one first heard the news of the 9/11 terrorist attacks (Brown & Kulik, 1977; Hirst et al., 2015). Flashbulb memories may be driven in part by the extreme novelty or surprise of the event – which would suggest a global effect of PE, rather than one specific to the cause of the PE – though it is clear that emotional factors and the significance of the event also play an important role. Indeed, there is evidence that surprise per se is not necessary for flashbulb memory (Coluccia et al., 2010).

In mundane laboratory studies, this incidental contextual information is often called “source information”, which can range from where on

⁴ Note that the flip case of this situation is that the memory for familiar items is better because of context surprise, as described in previous section, owing to the PE arising where the prior is flat but the likelihood is precise (Fig. 1C).

the computer screen an item was presented, to the mental thoughts that occurred at the same time (Johnson et al., 1993). For example, Kim et al. (2012) found that memory for source information (e.g., the location of an item or the background colour) was better for novel compared to repeated items (where repeated items were presented before the colour/location association in a pre-exposure phase)⁵. Likewise, Greve et al. (2019) found that memory for incidental information (in the sense that it was not task-relevant) was better when events violated the learned rule. In line with these findings, other studies report that low-frequency (and hence more surprising) words are associated with better source memory (i.e., word was spoken in a male or female voice) than high-frequency words (Guttentag & Carroll, 1994), or the mental operations performed (Guttentag & Carroll, 1997; Ye et al., 2019). On the other hand, some studies (e.g. Uncapher et al., 2011) report the opposite pattern of results, with worse source memory (for presentation side on the screen) for invalidly-cued than validly-cued trials in a Posner cuing task. It remains to be seen whether other factors beyond novelty, e.g. changes in attention, could explain these contradicting findings.

7.2. Temporal extent of the effect

The previous section considered the effect of novelty on simultaneous source information. However, what about information experienced shortly before or after a novel event? There is evidence that novelty can exert a beneficial effect both retroactively and proactively, so as to “spill-over” to stimuli that are not novel themselves, but are simply encountered in close temporal proximity. This effect, inspired by the synaptic tag and capture theory (Frey & Morris, 1997) was first studied extensively in animals (e.g., Ballarini et al., 2009; Moncada & Viola, 2007; Wang et al., 2010). For example, it is well documented that when a rodent is placed in a novel environment, memory for information learned immediately before or after that event is boosted (Wang et al., 2010). Similar effects have since been reported in humans (Ballarini et al., 2013; Bunzeck & Düzel, 2006; Fenker et al., 2008; Ramirez Butavand et al., 2020; Schomaker, van Bronkhorst, et al., 2014; but see Biel & Bunzeck, 2019). Fenker et al. (2008), for example, reported that participants who viewed novel versus familiar scenes showed better memory for unrelated words that were presented either before or after a scene (though see Biel & Bunzeck, 2019, for a failure to find this effect). A proactive memory benefit has also been reported by Schomaker, van Bronkhorst, et al. (2014) who employed immersive virtual reality and demonstrated enhanced memory for words that were learned immediately after exposure to a novel environment (analogous to the experiments on rodents). Similarly, Wittmann et al. (2007) reported superior memory for items that were preceded by a cue that predicted novelty, irrespective of whether the item itself was novel. Most impressively, the novelty effect on surrounding information has been tested in real-world educational setting by Ballarini et al. (2013), who demonstrated that teaching novel content to primary school children in an engaging science or music lesson, can boost memory for unrelated verbal and pictorial material studied up to one hour (but not four hours) before or after the lesson. This has been replicated in another set of high school students (Ramirez Butavand et al., 2020). These retroactive and proactive effects suggest that novelty causes short-term changes in general synaptic plasticity (see below), which facilitate consolidation processes (at least in the case of retroactive effects), in addition to the encoding processes we have considered so far. Moreover, the findings from Schomaker et al. (2014) and Fenker et al. (2008) were evident in recall and recollection which is in line with the idea that novelty enhances memory by engaging hippocampal encoding/consolidation processes (Kafkas & Montaldi, 2018b).

⁵ Note this finding appears to contradict that of the Kaula and Henson (2020) study described earlier; for possible reasons, see Discussion section of Kaula and Henson (2020).

8. Novelty and event boundaries

In addition to how well we remember information, novelty and surprise also appear to play an important role in parsing or segmenting our continuous experience. In contrast to most laboratory experiments that show discrete events/items/trials, our experiences in the real world are derived from a continuous stream of sensory input, during which our brains are constantly trying to predict what will be perceived next. According to “event segmentation theory” (Reynolds et al., 2007; Zacks et al., 2007), sudden changes in the input stream can evoke a violation of such predictions, which is believed to create boundaries that mark the beginning and the end of an event. In other words, prediction errors also act as a gating mechanism that parses our continuous perception into discrete memories.

This notion is supported by Rouhani et al. (2019), who associated scenes with a monetary reward and found that memory was worse for associations that spanned across high reward prediction error (RPE) trials than low RPE trials. This is in line with a range of studies that confirm memory is better for items that belong to the same event (DuBrow & Davachi, 2013, 2016; Ezzyat & Davachi, 2014; Heusser et al., 2018; Horner et al., 2016). Memory for event boundaries themselves has been reported to be superior (Rouhani et al., 2019; Swallow et al., 2009) and objects encountered at boundaries show stronger item-context binding (for images on colour backgrounds) compared to non-boundary objects (Heusser et al., 2018). This strengthens the view that an abrupt context shift in our environment can elicit a PE, which leads to boundaries that influence the formation, strength and structure of how we remember events. Indeed, neuroimaging has revealed peaks in activity in the hippocampus at event boundaries (Ben-Yakov & Henson, 2018), and the level of this “event offset” activity seems to correlate with how well the preceding event is remembered (Ben-Yakov & Dudai, 2011). Animal research suggests that such “spill-over” effects last a fixed period of time (see Okuda et al., 2020), so an interesting question for human studies is whether “spill-over” effects extend only as far as the surrounding event boundaries (Ben-Yakov et al., 2020).

9. Brain systems of novelty and surprise

So far, we have talked about the behavioural predictions of PIMMS. There is also a wealth of neuroimaging studies on humans (e.g., using PET, fMRI, EEG) that address the brain regions involved and/or the temporal dynamics of novelty effects, and an even larger literature on animal studies of the neurobiology of novelty and memory. We only scratch the surface of these literatures here, in an attempt to bring out the clearest findings in relation to some of the distinctions outlined in the previous sections.

9.1. Neuroimaging studies of novelty and surprise in humans

Context surprise has consistently been associated with the medial temporal lobe (MTL), and the hippocampus in particular, across a wide range of neuroimaging studies (e.g. Axmacher et al., 2010; Chen et al., 2015; Daselaar et al., 2006; Dolan & Fletcher, 1997; Grunwald et al., 1998; Grunwald & Lehnertz, 2003; Guitart-Masip et al., 2010; Knight, 1996; Köhler et al., 2005; Maass et al., 2014; Poppenk et al., 2008; Rutishauser et al., 2006; Tulving et al., 1996). Stronger hippocampal activity for context surprise was first reported in an early PET study by Tulving et al. (1996) who presented novel items intermixed with familiar items (see Fig. 2A). Concerns that this truly reflects activation for novel events rather than habituation to familiar events were overcome in later studies that tested “match-mismatch” paradigms (Kumaran & Maguire, 2006, 2007, 2009). These paradigms contrasted repeated sequences of objects (e.g., ABCD) with partially rearranged sequenced in which only the first two objects remaining intact (e.g., ABDC) and fully scrambled sequences (e.g., DACB; see Fig. 2G). Compared to repeated sequences (i.e., maximal habituation), or fully

scrambled sequences (i.e., maximal novelty), hippocampal activity was greater for partially rearranged sequences, for which predictions generated by the first two objects (AB) were violated by object D which occurred instead of C (i.e., maximal PE). This reaffirmed that the hippocampus tracks context surprise/PE, rather than indexing novel arrangements per se. Various other studies, assessing real-world predictions (Kang et al., 2009; Pine et al., 2018; Schiffer et al., 2012; Straube et al., 2014) and RPE (Davidow et al., 2016; though see Wimmer et al., 2014) corroborated the conclusion that the hippocampus processes contextual surprise. Importantly, not all of these studies related the novelty effects to subsequent memory, though those that did (e.g. Chen et al., 2015; Kirchhoff et al., 2000) generally find the same regions that show novelty/surprise, also predict subsequent memory.

These findings are in line with the PIMMS framework, which predicts strong hippocampal activity for context surprise (Fig. 1A), and that this should predict subsequent episodic memory. It should be noted, though, that empirical studies have also linked additional brain regions to contextual surprise, most prominently the locus coeruleus (LC), the substantia nigra, striatum and ventral tegmental area (Clos et al., 2019; Fenker et al., 2008; Guitart-Masip et al., 2010; Hollerman & Schultz, 1998; Kamiński et al., 2018; Mikell et al., 2014; Murty et al., 2016; Schott et al., 2004; Wittmann et al., 2007).

Studies that use event-related potentials (ERPs) to examine context surprise, report deflections of a central-parietal ERP component known as the P3, elicited 300–400 ms after stimulus onset (e.g. Friedman et al., 2001; Polich, 2007). Stronger P3 modulation has been reported for deviant relative to standard items in oddball paradigms (Nieuwenhuis et al., 2011; Polich, 2007). Notably, the P3 amplitude is sensitive to the difference in frequency of occurrence and level of deviation between standard and deviant items (Ranganath & Rainer, 2003; Nieuwenhuis et al., 2011), which endorses the view that this component reflects expectancy but not novelty per se. However, it is also important to relate these ERP novelty effects to memory encoding and here the current literature does not offer a coherent picture. While some studies show the P3 amplitude is linked to subsequent recall (Butterfield & Mangels, 2003; Fabiani et al., 1990; Fabiani & Donchin, 1995; Kamp et al., 2013; but see Rangel-Gomez & Meeter, 2013), others failed to observe P3 subsequent memory effects in recognition paradigms (Fabiani & Donchin, 1995). This discrepancy might be explained by the different types of memory tests, which are subserved by different processes. However, hippocampal intracranial EEG shows a component similar to the P3 that is related to subsequent memory effects (Axmacher et al., 2010), lending support for a general relationship between P3, hippocampus and memory function (for a review see Fonken et al., 2020).

In addition to context surprise, PIMMS predicts hippocampal activity for context novelty or item novelty, albeit lower, since it is driven by a smaller PE than for context surprise (Fig. 1). This has received some empirical support from intracranial ERP studies that show evoked potentials from hippocampus (N300/N350 and P600-like responses) are less pronounced for nonsense objects (item novelty) than real objects (Vannucci et al., 2003, 2008) in a context that holds no strong predictions. Interestingly, a difference between real objects and nonsense objects was only observed for patients with intact but not with impaired visual memory (Vannucci et al., 2008). However, some studies report the opposite finding, with stronger hippocampal activity for non-words than words, at least when tested with a lexical decision task (Graves et al., 2017; Mattheiss et al., 2018; though not when tested in a passive viewing condition, Braun et al., 2015). Further studies are needed to delineate whether task difficulty might confound the findings of item novelty in a lexical decision task because non-words might require more processing time than words (Graves et al., 2017).

Other ERP studies have examined item novelty. For example, a study by Cykowicz and Friedman (2007) reported a smaller P3 for repeated familiar symbols compared to repeated novel symbols, which was in turn associated with worse memory for the novel symbols (also see Cykowicz, 2019). This is also the case in tests using illegal non-words,

distorted pictures or unknown sounds (for reviews see Friedman et al., 2001; Rugg & Doyle, 1994). More importantly, the subsequent memory effect was larger for familiar than unfamiliar symbols (Cykowicz, 2019). These studies confirm that item novelty is associated with less effective encoding compared to familiar items. More generally, these ERP signatures suggest that it might be important to differentiate context surprise from item surprise, and possibly context/item surprise from context/item novelty (also see Schomaker, Roos, et al., 2014; Schomaker & Meeter, 2015).

9.2. Neurobiological models of novelty and surprise

So far, we sought to explain how novelty and surprise determine the fate of new declarative memories through the lens of PIMMS. Although PIMMS offers a framework for thinking about novelty and surprise, as well as making predictions for behaviour, it remains largely silent as to what brain mechanisms, neurotransmitters and molecular processes might underpin such behaviour. These mechanisms have been addressed by several neurobiological models of novelty and surprise in the animal literature. However, it is often difficult to reconcile those models with the data reviewed above, because of disparities in the terminology and nature of the paradigms used in the animal and human literature (see also Schomaker, 2019).

In line with human neuroimaging studies, most models propose that the hippocampus plays a central role, but make different claims about the neural circuits involved in processing novelty and surprise. Schomaker and Meeter (2015), for example, suggest stimulus and contextual novelty (or what we call item novelty and context novelty) engage the amygdala and enhance perceptual sensitivity, through the substantia nigra/ventral tegmental area (SN/VTA) and modulation of long term potentiation (LTP). More importantly, surprise, rather than novelty, is thought to engage a separate system including the anterior cingulate cortex (ACC), which activates the LC and norepinephrine system.

On the other hand, Duzkiewicz et al. (2019) propose that the hippocampus receives dopaminergic modulation from both the VTA and the locus coeruleus (LC). Indeed, the direct connection between hippocampus and LC might be even more critical than the indirect link to the VTA (Takeuchi et al., 2016). Duzkiewicz et al. (2019) distinguish “common novelty”, i.e. experiences that share some aspects with past experiences from “distinct novelty”, i.e. an experience that shares only few aspects with past experience (such as seeing the ocean for the first time). The latter is probably related closest to our term ‘complete novelty’, which is claimed to activate the LC and initiate consolidation processes in the hippocampus, which creates a memory trace associated with better recollection. Nevertheless, these data do not align well with PIMMS and the human literature, which shows rather poor encoding for item/complete novelty (see earlier), which might reflect inconsistencies in the translation of animal and human studies.

Interestingly, an influential framework (Düzel et al., 2010; Lisman et al., 2011; Lisman & Grace, 2005) suggests the hippocampus detects surprise (though referred to by the authors as “novelty”) of any type, and sends a signal to the SN/VTA via the striatum, which in turn releases dopamine in the hippocampus. Within the SN/VTA, these signals modulate phasic and tonic activity patterns. Düzel et al. (2010) speculated that phasic dopamine induces the production of plasticity-related products (PRPs, see below), while tonic dopamine activity increases the probability of phasic bursts, which lower threshold for learning. An important consequence of experiencing surprise according to this model is that it increases motivation and drives exploration (Düzel et al., 2010). Such effects are largely neglected in the human literature, but might play an important role and enhance memory encoding/consolidation besides PE.

Kafkas and Montaldi (2018b) propose that the anterior hippocampus is crucial for detecting surprise (referred to by the authors as “novelty”), at which point it sends a signal to other modulatory regions (e.g. midbrain and striatum). This includes release of acetylcholine, which

induces pupil constriction. Interestingly, Naber et al. (2013) also link pupil constriction during encoding of surprising items to better subsequent retrieval. This effect occurs regardless of whether an item is a target or non-target in the encoding task, although task relevance of targets further modulates pupil size (Kafkas & Montaldi, 2015a, 2015b; Otero et al., 2011). Importantly, other studies show the opposite finding, i.e., pupil dilation for surprise (Kafkas & Montaldi, 2015a; Kloosterman et al., 2015; Preuschoff et al., 2011) and curiosity (Kang et al., 2009). Pupil dilation is associated with the sympathetic nervous system, i.e., acetylcholine and noradrenaline, while pupil constriction is regulated by the parasympathetic nervous system, which engages acetylcholine only. These data seem to indicate that two distinct brain regions and neurotransmitter systems could be engaged by surprise, hence further differentiation of context surprise might be needed to distinguish surprise driven either by a rare occurrence of an item (i.e. Kafkas & Montaldi, 2015a) from surprise that is generated by expectations from previous presentations (i.e., Naber et al., 2013). More work is needed to distinguish these possibilities.

Many of the models mentioned so far base their assumptions on key findings and ideas in the animal literature, such as the “behavioural tagging theory” (BTT; Ballarini et al., 2009; Moncada & Viola, 2007). BTT itself is derived from the synaptic “tag-and-capture” theory (Frey & Morris, 1997; Redondo & Morris, 2011). This theory posits that the mechanism underlying the maintenance of late-stage LTP involves two processes. In the first step, a synapse is tagged as the result of recent input. After tagging, the synapse then captures so-called plasticity-related products (PRPs). This leads to the induction of lasting structural change in the synapse. A seminal experiment by Frey and Morris (1997) showed, when a synapse receives weak tetanisation, the synapse is tagged and early-LTP induced. However, without captured PRPs, early-LTP does not automatically become late-LTP, but strong tetanisation of a different synapse on the same neuronal population can do just that by providing PRPs to the weakly stimulated synapse. In this way, according to BTT, experiencing surprise (as novelty is operationalised in these studies) can have a similar effect as strong tetanisation and induce late-LTP after only weak stimulation (Li et al., 2003; Straube, Korz, & Frey, 2003; Straube, Korz, Balschun, et al., 2003).

Interestingly, a tag is believed to last approximately 90 min (Redondo & Morris, 2011), which means there is a critical time window during which a strong tetanisation (e.g., from novelty) has to be experienced to cause long-lasting memory change (Ballarini et al., 2009; Moncada & Viola, 2007). The time window is further influenced by other task characteristics and might follow a nonlinear function (Moncada et al., 2015). More importantly, this mechanism of time limited PRPs might explain the novelty/surprise effects for temporally-surrounding information discussed earlier in the human literature (see section Temporal extent of the effect). The human studies, however, show that the surprise-induced memory advantage for surrounding information can sometimes occur within a few seconds or minutes of the learning experience (e.g. Bunzeck & Düzel, 2006; Schomaker, van Bronkhorst, et al., 2014), which is too fast for the consolidation mechanisms involving PRPs to unfold. Future studies are needed to pinpoint the precise neurobiological mechanism underlying these effects in humans.

Finally, the interplay between PE, protein synthesis and consolidation is also captured in the phenomenon of reconsolidation, by which reactivated memory traces are destabilized and altered (Hardt, Einarsson, & Nader, 2010). This process depends on behavioural tagging (Rabinovich Orlandi et al., 2020), and is thought to require surprising events, i.e., PE, in order for reconsolidation to be effective. While this has been shown for non-declarative fear conditioning (Sevenster et al. (2013) and Sinclair & Barense (2018) propose that it can occur in human declarative memory too. They showed that video clips depicting action-outcomes are associated with higher rates of memory intrusions when the videos were surprisingly cut-off before the action, which they attributed to surprise that allowed memory traces to be modified by

reconsolidation.

10. How does PIMMS fare?

While PIMMS can explain some of the situations where surprise improves declarative memory, and some situations where simple novelty does not (or even produces worse memory), it clearly has several limitations. Foremost, PIMMS is a framework for thinking about different types of novelty (perhaps at Marr’s “algorithmic” level; Marr, 1971), and while tentative steps have been made to relate some of the layers in the assumed hierarchy to certain brain regions, it is not a theory about the underlying neuronal mechanisms (i.e., at Marr’s “implementation” levels). As reviewed in the above section on neurobiology, there are many complex processes involved in LTP, consolidation and various neurotransmitters that are also potentially affected by novelty, but about which PIMMS is silent. Moreover, though we have currently focused on just three levels in the hierarchy (the episodic, semantic and perceptual), in order to simplify and make contact with most of the human behavioural work on declarative memory, there are likely to be many more levels needed, for example to explain associative novelty, in terms of predictions between two objects that have been paired repeatedly. There are often also “meta-predictions” operating, such as whether or not surprise is itself expected. This has been studied in the context of volatile versus stable environments (Yu & Dayan, 2005), where a surprising event in a volatile (ever-changing) environment is less surprising than one occurring in a stable environment. A further consideration that is not fully addressed here is the role of attention in boosting memory for novel or surprising events. While attention may be a necessary mediator for improving memory, we do not think it is a sufficient explanation, because something else (e.g., PE) first has to cause a change in attention.

More importantly, there are also situations, mentioned above, where the behavioural evidence does not conform to PIMMS’s predictions. These include events that are highly congruent with expectations (or “schemas”; for review see Ghosh & Gilboa, 2014), and so elicit minimal PE and should not be encoded, and yet can still be recalled better (even when controlling for the benefits of schema at retrieval). This led to the proposal of a second brain system specialised for rapid consolidation of schema-congruent events (Alba & Hasher, 1983; Atienza et al., 2011; Cycowicz et al., 2008; Liu et al., 2018). Another situation that is problematic for PIMMS is when unknown objects are encountered in a novel environment that provides few expectations (i.e., flat prior and likelihood), such as a rodent entering a novel arena with novel objects (Tse et al., 2007), where research has demonstrated that this experience boosts memory for surrounding information. Not only does this situation produce negligible PE according to PIMMS, but PE-driven learning in PIMMS is assumed to occur at the time of encoding, so it is silent on processes (like those assumed by BTT) that improve memory for information occurring earlier or later in time, or on memory improvements that only emerge after a period of consolidation. While these cases could be used to indicate that PIMMS is not a good model of how novelty affects memory, we prefer to use these cases to argue that PE-driven memory encoding is not sufficient to explain the effects of all types of novelty, and those situations where PE-driven memory encoding cannot explain the data may be precisely those situations in which the brain needs (i.e., has evolved) other mechanisms by which novelty can influence memory. Clearly, there is an important challenge to develop unified theories that can explain the full and diverse range of ways in which novelty and memory interact.

Funding statement or declaration of conflicting interests

This work was supported by the United Kingdom Medical Research Council (SUAG/046 G101400). The Gates Cambridge Trust funds JAQ’s PhD studentship. The authors declare no conflicting interests.

CRediT authorship contribution statement

Jörn Alexander Quent: Conceptualization, Visualization, Writing - original draft, Writing - review & editing. **Richard N. Henson:** Conceptualization, Visualization, Writing - original draft, Writing - review & editing. **Andrea Greve:** Conceptualization, Visualization, Writing - original draft, Writing - review & editing.

References

- Åberg, C. S., & Nilsson, L. (2001). Facilitation of source discrimination in the novelty effect. *Scandinavian Journal of Psychology*, 42(4), 349–357. <https://doi.org/10.1111/1467-9450.00246>.
- Åberg, C. S., & Nilsson, L. G. (2003). A strict response criterion yields a mirror effect in the novelty paradigm. *Scandinavian Journal of Psychology*, 44(5), 425–432. <https://doi.org/10.1046/j.1467-9450.2003.00363.x>.
- Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal–anterior thalamic axis. *Behavioral and Brain Sciences*, 22(3), 425–444. <https://doi.org/10.1017/S0140525X99002034>.
- Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin*, 93(2), 203–231. <https://doi.org/10.1037/0033-2909.93.2.203>.
- Atienza, M., Crespo-García, M., & Cantero, J. L. (2011). Semantic congruence enhances memory of episodic associations: Role of theta oscillations. *Journal of Cognitive Neuroscience*, 23(1), 75–90. <https://doi.org/10.1162/jocn.2009.21358>.
- Axmacher, N., Cohen, M. X., Fell, J., Haupt, S., Dümpelmann, M., Elger, C. E., ... Ranganath, C. (2010). Intracranial EEG correlates of expectancy and memory formation in the human hippocampus and nucleus accumbens. *Neuron*, 65(4), 541–549. <https://doi.org/10.1016/j.neuron.2010.02.006>.
- Ballarín, F., Martínez, M. C., Díaz Pérez, M., Moncada, D., & Viola, H. (2013). Memory in elementary school children is improved by an unrelated novel experience. *PLoS ONE*, 8(6). <https://doi.org/10.1371/journal.pone.0066875>.
- Ballarín, F., Moncada, D., Martínez, M. C., Alen, N., & Viola, H. (2009). Behavioral tagging is a general mechanism of long-term memory formation. *Proceedings of the National Academy of Sciences*, 106(34), 14599–14604. <https://doi.org/10.1073/pnas.0907078106>.
- Belleville, S., Ménard, M. C., & Lepage, É. (2011). Impact of novelty and type of material on recognition in healthy older adults and persons with mild cognitive impairment. *Neuropsychologia*, 49(10), 2856–2865. <https://doi.org/10.1016/j.neuropsychologia.2011.06.011>.
- Ben-Yakov, A., & Dudai, Y. (2011). Constructing realistic engrams: Poststimulus activity of hippocampus and dorsal striatum predicts subsequent episodic memory. *Journal of Neuroscience*, 31(24), 9032–9042. <https://doi.org/10.1523/JNEUROSCI.0702-11.2011>.
- Ben-Yakov, A., Henson, R. N. (2018). The hippocampal film editor: Sensitivity and specificity to event boundaries in continuous experience. *Journal of Neuroscience*, 38(47), 10057–10068. <https://doi.org/10.1523/JNEUROSCI.0524-18.2018>.
- Ben-Yakov, A., Smith, V., & Henson, R. N. (2020). The temporal dominion of surprise. *Psychonomic Bulletin & Review*.
- Berlyne, D. E. (1960). In *Conflict, arousal, and curiosity*. McGraw-Hill Book Company. <https://doi.org/10.1037/11164-000>.
- 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>.
- Braun, M., Jacobs, A. M., Richlan, F., Hawelka, S., Hutzler, F., & Kronbichler, M. (2015). Many neighbors are not silent. fMRI evidence for global lexical activity in visual word recognition. *Frontiers in Human Neuroscience*, 9. <https://doi.org/10.3389/fnhum.2015.00423>.
- Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, 5(1), 73–99. [https://doi.org/10.1016/0010-0277\(77\)90018-X](https://doi.org/10.1016/0010-0277(77)90018-X).
- Bunzeck, N., & Düzel, E. (2006). Absolute coding of stimulus novelty in the human substantia nigra/STN. *Neuron*, 51(3), 369–379. <https://doi.org/10.1016/j.neuron.2006.06.021>.
- Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research*, 17(3), 793–817. [https://doi.org/10.1016/S0926-6410\(03\)00203-9](https://doi.org/10.1016/S0926-6410(03)00203-9).
- Calderon, C. B., Loof, E. De, Ergo, K., Snoeck, A., Boehler, C. N., & Verguts, T. (2020). Signed reward prediction errors in the ventral striatum drive episodic memory. *bioRxiv*, 0, 2020.01.03.893578. <https://doi.org/10.1101/2020.01.03.893578>.
- Chen, J., Cook, P. A., & Wagner, A. D. (2015). Prediction strength modulates responses in human area CA1 to sequence violations. *Journal of Neurophysiology*, 114(2), 1227–1238. <https://doi.org/10.1152/jn.00149.2015>.
- Clos, M., Bunzeck, N., & Sommer, T. (2019). Dopamine enhances item novelty detection via hippocampal and associative recall via left lateral prefrontal cortex mechanisms. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 39(40), 7920–7933. <https://doi.org/10.1523/JNEUROSCI.0495-19.2019>.
- Coluccia, E., Bianco, C., & Brandimonte, M. A. (2010). Autobiographical and event memories for surprising and unsurprising events. *Applied Cognitive Psychology*, 24(2), 177–199. <https://doi.org/10.1002/acp.1549>.
- Cycowicz, Y. M. (2019). Orienting and memory to unexpected and/or unfamiliar visual events in children and adults. *Developmental Cognitive Neuroscience*, 36, Article 100615. <https://doi.org/10.1016/j.dcn.2019.100615>.
- Cycowicz, Y. M., & Friedman, D. (2007). Visual novel stimuli in an ERP novelty oddball paradigm: Effects of familiarity on repetition and recognition memory. *Psychophysiology*, 44(1), 11–29. <https://doi.org/10.1111/j.1469-8986.2006.00481.x>.
- Cycowicz, Y. M., Nessler, D., Horton, C., & Friedman, D. (2008). Retrieving object color: The influence of color congruity and test format. *NeuroReport*, 19(14), 1387–1390. <https://doi.org/10.1097/WNR.0b013e32830c8df1>.
- Daselaar, S. M., Fleck, M. S., & Cabeza, R. (2006). Triple dissociation in the medial temporal lobes: Recollection, familiarity, and novelty. *Journal of Neurophysiology*, 96(4), 1902–1911. <https://doi.org/10.1152/jn.01029.2005>.
- Davidow, J. Y., Foerster, K., Galván, A., & Shohamy, D. (2016). An upside to reward sensitivity: The hippocampus supports enhanced reinforcement learning in adolescence. *Neuron*, 92(1), 93–99. <https://doi.org/10.1016/j.neuron.2016.08.031>.
- De Loof, E., Ergo, K., Naert, L., Janssens, C., Talsma, D., Van Opstal, F., & Verguts, T. (2018). Signed reward prediction errors drive declarative learning. *PLoS ONE*, 13(1), Article e0189212. <https://doi.org/10.1371/journal.pone.0189212>.
- Dobbins, I. G., Kroll, N. E. A., Yonelinas, A. P., & Liu, Q. (1998). Distinctiveness in recognition and free recall: The role of recollection in the rejection of the familiar. *Journal of Memory and Language*, 38(4), 381–400. <https://doi.org/10.1006/jmla.1997.2554>.
- Dolan, R. J., & Fletcher, P. C. (1997). Dissociating prefrontal and hippocampal function in episodic memory encoding. *Nature*, 388(6642), 582–585. <https://doi.org/10.1038/41561>.
- DuBrow, S., & Davachi, L. (2013). The influence of context boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, 142(4), 1277–1286. <https://doi.org/10.1037/a0034024>.
- DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory*, 134, 107–114. <https://doi.org/10.1016/j.nlm.2016.07.011>.
- 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>.
- Düzel, E., Bunzeck, N., Guitart-Masip, M., & Düzel, S. (2010). Novelty-related Motivation of Anticipation and exploration by Dopamine (NOMAD): Implications for healthy aging. *Neuroscience and Biobehavioral Reviews*, 34(5), 660–669. <https://doi.org/10.1016/j.neubiorev.2009.08.006>.
- Elliott Wimmer, G., Braun, E. K., Daw, N. D., & Shohamy, D. (2014). Episodic memory encoding interferes with reward learning and decreases striatal prediction errors. *Journal of Neuroscience*, 34(45), 14901–14912. <https://doi.org/10.1523/JNEUROSCI.0204-14.2014>.
- Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, 81(5), 1179–1189. <https://doi.org/10.1016/j.neuron.2014.01.042>.
- Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(1), 224–240. <https://doi.org/10.1037/0278-7393.21.1.224>.
- Fabiani, M., Gratton, G., Chiarenza, G., & Donchin, E. (1990). A psychophysiological investigation of the von Restorff paradigm in children. *Journal of Psychophysiology*, 4(1), 15–24.
- Fenker, D. B., Frey, J. U., Schuetz, 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. <https://doi.org/10.1162/jocn.2008.20086>.
- 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>.
- Foster, M. L., & Keane, M. T. (2019). The role of surprise in learning: different surprising outcomes affect memorability differentially. *Topics in Cognitive Science*, 11(1), 75–87. <https://doi.org/10.1111/tops.12392>.
- Frey, U., & Morris, R. G. M. (1997). Synaptic tagging and long-term potentiation. *Nature*, 385(6616), 533–536. <https://doi.org/10.1038/385533a0>.
- Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: An event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews*, 25(4), 355–373. [https://doi.org/10.1016/S0149-7634\(01\)00019-7](https://doi.org/10.1016/S0149-7634(01)00019-7).
- Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, 18(1), 23–30. <https://doi.org/10.3758/BF03202642>.
- Ghosh, V. E., & Gilboa, A. (2014). What is a memory schema? A historical perspective on current neuroscience literature. *Neuropsychologia*, 53, 104–114. <https://doi.org/10.1016/j.neuropsychologia.2013.11.010>.
- Graves, W. W., Boukrina, O., Mattheiss, S. R., Alexander, E. J., & Baillet, S. (2017). Reversing the standard neural signature of the word-nonword distinction. *Journal of Cognitive Neuroscience*, 29(1), 79–94. https://doi.org/10.1162/jocn_a.01022.
- Greene, R. L. (1999). The role of familiarity in recognition. *Psychonomic Bulletin & Review*, 6(2), 309–312. <https://doi.org/10.3758/BF03212335>.
- Greve, A., Cooper, E., Kaula, A., Anderson, M. C., & Henson, R. (2017). Does prediction error drive one-shot declarative learning? *Journal of Memory and Language*, 94, 149–165. <https://doi.org/10.1016/j.jml.2016.11.001>.
- 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>.
- Gruber, M. J., & Ranganath, C. (2019). How curiosity enhances hippocampus-dependent memory: The prediction, appraisal, curiosity, and exploration (PACE) framework. *Trends in Cognitive Sciences*, 23(12), 1014–1025. <https://doi.org/10.1016/j.tics.2019.10.003>.

- Grunwald, T., & Lehnertz, K. (2003). Novelty detection and encoding for memory within the human hippocampal system. *International IEEE/EMBS Conference on Neural Engineering, NER*, 2003-Janua(4), 32–34. <https://doi.org/10.1109/CNE.2003.1196748>.
- Grunwald, T., Lehnertz, K., Heinze, H. J., Helmstaedter, C., & Elger, C. E. (1998). Verbal novelty detection within the human hippocampus proper. *Proceedings of the National Academy of Sciences of the United States of America*, 95(6), 3193–3197. <https://doi.org/10.1073/pnas.95.6.3193>.
- Guitart-Masip, M., Bunzeck, N., Stephan, K. E., Dolan, R. J., & Düzel, E. (2010). Contextual novelty changes reward representations in the striatum. *Journal of Neuroscience*, 30(5), 1721–1726. <https://doi.org/10.1523/JNEUROSCI.5331-09.2010>.
- Guttenberg, R. E., & Carroll, D. (1994). Identifying the basis for the word frequency effect in recognition memory. *Memory*, 2(3), 255–273. <https://doi.org/10.1080/09658219408258948>.
- Guttenberg, R. E., & Carroll, D. (1997). Recollection-based recognition: Word frequency effects. *Journal of Memory and Language*, 37(4), 502–516. <https://doi.org/10.1006/jmla.1997.2532>.
- Hardt, O., Einarsson, E. Ö., & Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual Review of Psychology*, 61(1), 141–167. <https://doi.org/10.1146/annurev.psych.093008.100455>.
- Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, 20(11), 1315–1326. <https://doi.org/10.1002/hipo.20857>.
- Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual boundaries cause mnemonic trade-offs between local boundary processing and across-trial associative binding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(7), 1075–1090. <https://doi.org/10.1037/xlm0000503>.
- Hirst, W. M., Johnson, R., Buckner, M. K., Gabrieli, R. L., Mather, J. D. E., Schacter, M., D. Lyle, K. B., Phelps, E. A., C.J., V., Mitchell, K. J., Budson, A. E., Lustig, C., Ochsner, K. N., & Simons, J.S. Cuc, A. F. and O. A. (2015). A ten-year follow-up of a study of memory for the attack of September 11, 2...: EBSCOhost. *Journal of Experimental Psychology: General*, 144(3), 604–623.
- Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, 1(4), 304–309. <https://doi.org/10.1038/1124>.
- Hornet, A. J., Bisby, J. A., Wang, A., Bogus, K., & Burgess, N. (2016). The role of spatial boundaries in shaping long-term event representations. *Cognition*, 154, 151–164. <https://doi.org/10.1016/j.cognition.2016.05.013>.
- Jang, A. I., Nassar, M. R., Dillon, D. G., & Frank, M. J. (2019). Positive reward prediction errors during decision-making strengthen memory encoding. *Nature Human Behaviour*, 3(7), 719–732. <https://doi.org/10.1038/s41562-019-0597-3>.
- Java, R. I., Kaminska, Z., & Gardiner, J. M. (1995). Recognition Memory and Awareness for Famous and Obscure Musical Themes. *European Journal of Cognitive Psychology*, 7(1), 41–53. <https://doi.org/10.1080/09541449508520156>.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3–28. <https://doi.org/10.1037/0033-2909.114.1.3>.
- 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(December 2017), 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(July 2017), 60–68. <https://doi.org/10.1016/j.neulet.2018.01.053>.
- Kaminiski, J., Mamelak, A. N., Birch, K., Mosher, C. P., Tagliati, M., & Rutishauser, U. (2018). Novelty-Sensitive Dopaminergic Neurons in the Human Substantia Nigra Predict Success of Declarative Memory Formation. *Current Biology*, 28(9), 1333–1343.e4. <https://doi.org/10.1016/j.cub.2018.03.024>.
- Kamp, S.-M., Brumback, T., & Donchin, E. (2013). The component structure of ERP subsequent memory effects in the Von Restorff paradigm and the word frequency effect in recall. *Psychophysiology*, 50(11), 1079–1093. <https://doi.org/10.1111/psyp.12090>.
- Kang, M. J., Hsu, M., Krajchich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T., & Camerer, C. F. (2009). The wick in the candle of learning: Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science*, 20(8), 963–973. <https://doi.org/10.1111/j.1467-9280.2009.02402.x>.
- Kaula, A. J., & Henson, R. N. (2020). Priming effects on subsequent episodic memory: Testing attentional accounts. *Journal of Memory and Language*, 113, Article 104106. <https://doi.org/10.1016/j.jml.2020.104106>.
- Kim, K., Yi, D. J., Raye, C. L., & Johnson, M. K. (2012). Negative effects of item repetition on source memory. *Memory and Cognition*, 40(6), 889–901. <https://doi.org/10.3758/s13421-012-0196-2>.
- Kinsbourne, M., & George, J. (1974). The mechanism of the word-frequency effect on recognition memory. *Journal of Verbal Learning and Verbal Behavior*, 13(1), 63–69. [https://doi.org/10.1016/S0022-5371\(74\)80031-9](https://doi.org/10.1016/S0022-5371(74)80031-9).
- Kirchhoff, B. A., Wagner, A. D., Maril, A., & Stern, C. E. (2000). Prefrontal-Temporal Circuitry for Episodic Encoding and Subsequent Memory. *The Journal of Neuroscience*, 20(16), 6173 LP-6180. <https://doi.org/10.1523/JNEUROSCI.20-16-06173.2000>.
- Kishiyama, M. M., & Yonelinas, A. P. (2003). Novelty effects on recollection and familiarity in recognition memory. *Memory & Cognition*, 31(7), 1045–1051. <https://doi.org/10.3758/BF03196125>.
- Kloosterman, N. A., Meindertsma, T., van Loon, A. M., Lamme, V. A. F., Bonne, Y. S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise. *European Journal of Neuroscience*, 41(8), 1068–1078. <https://doi.org/10.1111/ejn.12859>.
- Knight, R. T. (1996). Novelty Detection Model. *Letters to Nature*, 383(September), 256–259.
- Köhler, S., Danckert, S., Gati, J. S., & Menon, R. S. (2005). Novelty responses to relational and non-relational information in the hippocampus and the parahippocampal region: A comparison based on event-related fMRI. *Hippocampus*, 15(6), 763–774. <https://doi.org/10.1002/hipo.20098>.
- Kormi-Nouri, R., Nilsson, L. G., & Ohta, N. (2005). The novelty effect: Support for the novelty-encoding hypothesis. *Scandinavian Journal of Psychology*, 46(2), 133–143. <https://doi.org/10.1111/j.1467-9450.2005.00443.x>.
- Kroll, J. F., & Potter, M. C. (1984). Recognizing words, pictures, and concepts: A comparison of lexical, object, and reality decisions. *Journal of Verbal Learning and Verbal Behavior*, 23(1), 39–66. [https://doi.org/10.1016/S0022-5371\(84\)90499-7](https://doi.org/10.1016/S0022-5371(84)90499-7).
- Kumaran, D., & Maguire, E. A. (2006). An unexpected sequence of events: Mismatch detection in the human hippocampus. *PLoS Biology*, 4(12), Article e424. <https://doi.org/10.1371/journal.pbio.0040424>.
- Kumaran, D., & Maguire, E. A. (2007). Which computational mechanisms operate in the hippocampus during novelty detection? *Hippocampus*, 17(9), 735–748. <https://doi.org/10.1002/hipo.20326>.
- Kumaran, D., & Maguire, E. A. (2009). Novelty signals: A window into hippocampal information processing. *Trends in Cognitive Sciences*, 13(2), 47–54. <https://doi.org/10.1016/j.tics.2008.11.004>.
- Li, S., Cullen, W. K., Anwyl, R., & Rowan, M. J. (2003). Dopamine-dependent facilitation of LTP induction in hippocampal CA1 by exposure to spatial novelty. *Nature Neuroscience*, 6(5), 526–531.
- 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>.
- Lisman, J. E., Grace, A. A., & Düzel, E. (2011). A neoHebbian framework for episodic memory; role of dopamine-dependent late LTP. *Trends in Neurosciences*, 34(10), 536–547. <https://doi.org/10.1016/j.tics.2011.07.006>.
- Liu, Z.-X., Grady, C., & Moscovitch, M. (2018). The effect of prior knowledge on post-encoding brain connectivity and its relation to subsequent memory. *NeuroImage*, 167, 211–223. <https://doi.org/10.1016/j.neuroimage.2017.11.032>.
- Maass, S., Schütze, H., Speck, O., Yonelinas, A., Tempelmann, C., Heinze, H.-J., ... Düzel, E. (2014). Laminar activity in the hippocampus and entorhinal cortex related to novelty and episodic encoding. *Nature Communications*, 5, 5547. <https://doi.org/10.1038/ncomms5547>.
- Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, 87(3), 252–271. <https://doi.org/10.1037/0033-295X.87.3.252>.
- Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 262(841), 23–81. <https://doi.org/10.1098/rstb.1971.0078>.
- Mattheiss, S. R., Levinson, H., & Graves, W. W. (2018). Duality of function: Activation for meaningless nonwords and semantic codes in the same brain areas. *Cerebral Cortex*, 28(7), 2516–2524. <https://doi.org/10.1093/cercor/bhy053>.
- Mayes, A. R., Montaldi, D., & Migo, E. (2007). Associative memory and the medial temporal lobes. *Trends in Cognitive Sciences*, 11(3), 126–135. <https://doi.org/10.1016/j.tics.2006.12.003>.
- Mikell, C. B., Sheehy, J. P., Youngerman, B. E., McGovern, R. A., Wojtasiewicz, T. J., Chan, A. K., ... McKhann, G. M. (2014). Features and timing of the response of single neurons to novelty in the substantia nigra. *Brain Research*, 1542, 79–84. <https://doi.org/10.1016/j.brainres.2013.10.033>.
- Moncada, D., Ballardini, F., Martinez, M. C., & Viola, H. (2015). In The Behavioral Tagging Hypothesis and Its Implications for Long-Term Memory Formation BT - Synaptic Tagging and Capture: From Synapses to Behavior (pp. 231–259). New York: Springer. https://doi.org/10.1007/978-1-4939-1761-7_14.
- Moncada, D., & Viola, H. (2007). Induction of Long-Term Memory by Exposure to Novelty Requires Protein Synthesis: Evidence for a Behavioral Tagging. *Journal of Neuroscience*, 27(28), 7476–7481. <https://doi.org/10.1523/JNEUROSCI.1083-07.2007>.
- 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>.
- Morris, R. G. M. (2006). Elements of a neurobiological theory of hippocampal function: The role of synaptic plasticity, synaptic tagging and schemas. *European Journal of Neuroscience*, 23(11), 2829–2846. <https://doi.org/10.1111/j.1460-9568.2006.04888.x>.
- Moscovitch, M. (1995). Recovered consciousness: A hypothesis concerning modularity and episodic memory. *Journal of Clinical and Experimental Neuropsychology*, 17(2), 276–290. <https://doi.org/10.1080/01686869508405123>.
- Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic Memory and Beyond: The Hippocampus and Neocortex in Transformation. *Annual Review of Psychology*, 67, 105–134. <https://doi.org/10.1146/annurev-psych-113011-143733>.
- Murty, V. P., Ballard, I. C., & Adcock, R. A. (2016). Hippocampus and Prefrontal Cortex Predict Distinct Timescales of Activation in the Human Ventral Tegmental Area. *Cerebral Cortex*, bhw005. <https://doi.org/10.1093/cercor/bhw005>.
- Naber, M., 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), 1–20. <https://doi.org/10.1167/13.2.11>.

- Nieuwenhuis, S., De Geus, E. J., & Aston-Jones, G. (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology*, 48(2), 162–175. <https://doi.org/10.1111/j.1469-8986.2010.01057.x>.
- O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. *Science*, 304(5669), 452–454. <https://doi.org/10.1126/science.1094285>.
- Okuda, K., Høigaard, 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>.
- Perfect, T. J., & Dasgupta, Z. R. (1997). What underlies the deficit in reported recollective experience in old age? *Memory and Cognition*, 25(6), 849–858. <https://doi.org/10.3758/BF03211329>.
- Pine, A., Sadeh, N., Ben-Yakov, A., Dudai, Y., & Mendelsohn, A. (2018). Knowledge acquisition is governed by striatal prediction errors. *Nature Communications*, 9(1), 1–14. <https://doi.org/10.1038/s41467-018-03992-5>.
- 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>.
- Poppenk, J., Köhler, S., & Moscovitch, M. (2010). Revisiting the novelty effect: When familiarity, not novelty, enhances memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 36(5), 1321–1330. <https://doi.org/10.1037/a0019900>.
- Poppenk, J., Walia, G., McIntosh, A. R., Joanisse, M. F., Klein, D., & Köhler, S. (2008). Why is the meaning of a sentence better remembered than its form? An fMRI study on the role of novelty-encoding processes. *Hippocampus*, 18(9), 909–918. <https://doi.org/10.1002/hipo.20453>.
- Preuschhoff, K., 't Hart, B. M., & Einhäuser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in Neuroscience*, 5 (SEP), 1–12. <https://doi.org/10.3389/fnins.2011.00115>.
- Prull, M. W. (2015). Adult age differences in memory for schema-consistent and schema-inconsistent objects in a real-world setting. *Aging, Neuropsychology, and Cognition*, 22 (6), 731–754. <https://doi.org/10.1080/13825585.2015.1037821>.
- Quent, J. A., & Henson, R. N. A. (2018). Effect of schema expectancy on object/location memory in VR. <http://doi.org/10.17605/OSF.IO/S9ER3>.
- Rabinovich Orlandi, I., Fullio, C. L., Schroeder, M. N., Giurfa, M., Ballarini, F., & Moncada, D. (2020). Behavioral tagging underlies memory reconsolidation. *Proceedings of the National Academy of Sciences*, 117(30), 18029–18036. <https://doi.org/10.1073/pnas.2009517117>.
- Ramirez Butavand, D., Hirsch, I., Tomaiuolo, M., Moncada, D., Viola, H., & Ballarini, F. (2020). Novelty Improves the Formation and Persistence of Memory in a Naturalistic School Scenario. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00048>.
- Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience*, 4, 193. <https://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>.
- Redondo, R. L., & Morris, R. G. M. (2011). Making memories last: The synaptic tagging and capture hypothesis. *Nature Reviews Neuroscience*, 12(1), 17–30. <https://doi.org/10.1038/nrn2963>.
- Reggev, N., Sharoni, R., & Maril, A. (2018). Distinctiveness Benefits Novelty (and Not Familiarity), but Only Up to a Limit: The Prior Knowledge Perspective. *Cognitive Science*, 42(1), 103–128. <https://doi.org/10.1111/cogs.12498>.
- Reichardt, R., Polner, B., & Simor, P. (2020). Novelty Manipulations, Memory Performance, and Predictive Coding: The Role of Unexpectedness. *Frontiers in Human Neuroscience*, 14(April), 1–11. <https://doi.org/10.3389/fnhum.2020.00152>.
- Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, 31(4), 613–643. <https://doi.org/10.1080/15326900701399913>.
- Rouhani, N., & Niv, Y. (2019). Depressive symptoms bias the prediction-error enhancement of memory towards negative events in reinforcement learning. *Psychopharmacology (Berl)*, 236(8), 2425–2435. <https://doi.org/10.1007/s00213-019-05322-z>.
- Rouhani, N., Norman, K. A., & Niv, Y. (2018). Dissociable effects of surprising rewards on learning and memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 44(9), 1430–1443. <https://doi.org/10.1037/xlm0000518>.
- Rouhani, N., Norman, K. A., Niv, Y., & Bornstein, A. M. (2019). Reward prediction errors create event boundaries in memory. *bioRxiv*, Article 725440. <https://doi.org/10.1101/725440>.
- Rugg, M. D., & Doyle, M. C. (1994). Event-Related Potentials and Stimulus Repetition in Direct and Indirect Tests of Memory BT - Cognitive Electrophysiology (H.-J. Heinze, T. F. Münte, & G. R. Mangun (eds.); pp. 124–148). Birkhäuser Boston. https://doi.org/10.1007/978-1-4612-0283-7_5.
- Rutishauser, U., Mamelak, A. N., & Schuman, E. M. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron*, 49(6), 805–813. <https://doi.org/10.1016/j.neuron.2006.02.015>.
- Schiffer, A. M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS ONE*, 7(5). <https://doi.org/10.1371/journal.pone.0036445>.
- Schomaker, J. (2019). Unexplored territory: Beneficial effects of novelty on memory. *Neurobiology of Learning and Memory*. <https://doi.org/10.1016/j.nlm.2019.03.005>.
- Schomaker, J., & Meeter, M. (2015). Short- and long-lasting consequences of novelty, deviance and surprise on brain and cognition. *Neuroscience and Biobehavioral Reviews*, 55, 268–279. <https://doi.org/10.1016/j.neubiorev.2015.05.002>.
- Schomaker, J., Roos, R., & Meeter, M. (2014). Expecting the unexpected: The effects of deviance on novelty processing. *Behavioral Neuroscience*, 128(2), 146–160. <https://doi.org/10.1037/a0036289>.
- 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(AUG), 1–6. <https://doi.org/10.3389/fpsyg.2014.00918>.
- Schott, B. H., Sellner, D. B., Lauer, C.-J., Habib, R., Frey, J. U., Guderian, S., ... Düzel, E. (2004). Activation of Midbrain Structures by Associative Novelty and the Formation of Explicit Memory in Humans. *Learning & Memory*, 11(4), 383–387. <https://doi.org/10.1101/lm.75004>.
- Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery & Psychiatry*, 20(1), 11–21. <https://doi.org/10.1136/jnnp.20.1.11>.
- Sevenster, D., Beckers, T., & Kindt, M. (2013). Prediction Error Governs Pharmacologically Induced Amnesia for Learned Fear. *Science*, 339(6121), 830–833. <https://doi.org/10.1126/science.1231357>.
- Sinclair, A. H., & Barense, M. D. (2018). Surprise and Destabilize: Prediction Error Triggers Episodic Memory Updating. *Learning & Memory*, 25, 369–381. <https://doi.org/10.1101/lm.046912.117.25>.
- Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, 99(2), 195–231. <https://doi.org/10.1037/0033-295X.99.2.195>.
- Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, 82(3), 171–177. <https://doi.org/10.1016/j.nlm.2004.06.005>.
- Staresina, B. P., & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *Journal of Neuroscience*, 26(36), 9162–9172. <https://doi.org/10.1523/JNEUROSCI.2877-06.2006>.
- Straube, B., Meyer, L., Green, A., & Kircher, T. (2014). Semantic relation vs. surprise: The differential effects of related and unrelated co-verbal gestures on neural encoding and subsequent recognition. *Brain Research*, 1567, 42–56. <https://doi.org/10.1016/j.brainres.2014.04.012>.
- Straube, T., Korf, V., Balschun, D., & Frey, J. U. (2003). Requirement of β -adrenergic receptor activation and protein synthesis for LTP-reinforcement by novelty in rat dentate gyrus. *Journal of Physiology*, 552(3), 953–960. <https://doi.org/10.1113/jphysiol.2003.049452>.
- Straube, T., Korf, V., & Frey, J. U. (2003). Bidirectional modulation of long-term potentiation by novelty-exploration in rat dentate gyrus. *Neuroscience Letters*, 344(1), 5–8. [https://doi.org/10.1016/S0304-3940\(03\)00349-5](https://doi.org/10.1016/S0304-3940(03)00349-5).
- Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, 138(2), 236–257. <https://doi.org/10.1037/a0015631>.
- Takeuchi, T., Duszkievicz, 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>.
- Tse, D., Langston, R. R. F., Kakeyama, M., Bethus, I., Spooner, P. a, Wood, E. R., Witter, M. P., & Morris, R. G. M. (2007). Schemas and memory consolidation. *Science*, 316 (5821), 76–82. <https://doi.org/10.1126/science.1153935>.
- Tulving, E., & Kroll, N. (1995). Novelty assessment in the brain and long-term memory encoding. *Psychonomic Bulletin & Review*, 2(3), 387–390. <https://doi.org/10.3758/BF03210977>.
- Tulving, E., Markowitsch, H. J., Craik, F. I. M., Habib, R., & Houle, S. (1996). Novelty and Familiarity Activations in PET Studies of Memory Encoding and Retrieval. *Cerebral Cortex*, 6(1), 71–79. <https://doi.org/10.1093/cercor/6.1.71>.
- Uncapher, M. R., Benjamin Hutchinson, J., & Wagner, A. D. (2011). Dissociable effects of top-down and bottom-up attention during episodic encoding. *Journal of Neuroscience*, 31(35), 12613–12628. <https://doi.org/10.1523/JNEUROSCI.0152-11.2011>.
- van Kesteren, M. T. R., Ruiter, D. J., Fernández, G., & Henson, R. N. (2012). How schema and novelty augment memory formation. *Trends in Neurosciences*, 35(4), 211–219. <https://doi.org/10.1016/j.tins.2012.02.001>.
- Vannucci, M., Dietl, T., Pezer, N., Viggiano, M. P., Helmstaedter, C., Schaller, C., ... Grunwald, T. (2003). Hippocampal function and visual object processing in temporal lobe epilepsy. *NeuroReport*, 14(11), 1489–1492. <https://doi.org/10.1097/00001756-200308060-00017>.
- Vannucci, M., Pezer, N., Helmstaedter, C., Schaller, K., Viggiano, M. P., Elger, C. E., & Grunwald, T. (2008). Hippocampal response to visual objects is related to visual memory functioning. *NeuroReport*, 19(9), 965–968. <https://doi.org/10.1097/WNR.0b013e328328302c89c>.
- von Restorff, H. (1933). Analyse von Vorgängen im Spurenfeld. *Psychologische Forschung*, 61(1912), 299–342.
- Wang, S.-H., Redondo, R. L., & Morris, R. G. M. (2010). Relevance of synaptic tagging and capture to the persistence of long-term potentiation and everyday spatial memory. *Proceedings of the National Academy of Sciences*, 107(45), 19537–19542. <https://doi.org/10.1073/pnas.1008638107>.
- Watkins, M. J., & Gardiner, J. M. (1979). An appreciation of generate-recognize theory of recall. *Journal of Verbal Learning and Verbal Behavior*, 18(6), 687–704. [https://doi.org/10.1016/S0022-5371\(79\)90397-9](https://doi.org/10.1016/S0022-5371(79)90397-9).
- Wittmann, B. C., Bunzeck, N., Dolan, R. J., & Düzel, E. (2007). Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *NeuroImage*, 38(1), 194–202. <https://doi.org/10.1016/j.neuroimage.2007.06.038>.

- Ye, J., Nie, A., & Liu, S. (2019). How do word frequency and memory task influence directed forgetting: An ERP study. *International Journal of Psychophysiology*, 146 (October), 157–172. <https://doi.org/10.1016/j.ijpsycho.2019.10.005>.
- 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>.
- Yu, A. J., & Dayan, P. (2005). Uncertainty, Neuromodulation, and Attention. *Neuron*, 46 (4), 681–692. <https://doi.org/10.1016/j.neuron.2005.04.026>.
- Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, 133(2), 273–293. <https://doi.org/10.1037/0033-2909.133.2.273>.