Surprise, Emotion and Memory: The interaction of surprising feedback and emotional valence within an image recognition task

Lederer, Moritz

Universiteit Leiden
Psychologie
Faculteit der Sociale Wetenschappen

Surprise, Emotion and Memory

The interaction of surprising feedback and emotional valence within an image recognition task

Name: Moritz Lederer
Date: 23.12.2021
Supervisor: Dr. David Amadeus Vogelsang
Second reader: Dr. Samarth Varma
Word count: 10727
Cognitive Psychology
Thesis Msci Applied Cognitive Psychology

Abstract

Even though surprise is a common emotion or experience in everyday life, we rarely think about the effects it can have on us. Previous research has shown that a surprise enhances memory, but not much is known about the cognitive mechanisms underlying this effect. Therefore, the current research investigated whether different strengths of surprise (mild or strong) or different directions of surprise (positive or negative) differentially influence memory. Based on the literature it was hypothesized that positive rather than negative and stronger rather than weaker surprises are better remembered. Moreover, it was investigated whether the emotional valence (positive or negative) of an image has an effect on memory. Participants (N = 25) undertook a recognition memory experiment with emotionally valent images spanning over three phases. Thereby participants encountered surprises of different strengths and of different directions. Based on the participants responses it was determined which level of surprise had the biggest effect on memorization and whether emotional valence had an effect. The results showed a significant interaction of the strength and direction of surprise. The interaction indicated that images associated with no surprise and images associated with strong negative surprise were significantly better remembered than images associated with a strong positive or a mild negative surprise. Furthermore, it was found that negative images are more frequently correctly recognized than positive images. Overall, the results of the study showed that only a strong negative surprise enhances memory for associated images but that also no surprise at all led to better memory. These findings can be of great relevance in education and learning research.

The interaction of surprising feedback and emotional valence in an image recognition task

Surely most of us are familiar with the phenomenon that memories for very salient events in our life are recalled more vividly and in much more detail. These so-called flashbulb memories can occur for very negative events like a relative's death but also for very positive events like finally expecting a child (Kraha & Boals, 2014). Two crucial factors determining the occurrence and strength of flashbulb memories are emotion and surprise (Conway et al., 1994; Kraha & Boals, 2014). That is, mostly the events that come by surprise and are highly emotionally significant to us are the ones remembered most vividly.

However, apart from these extreme events, surprise and emotion were also shown to enhance memory in everyday life. Surprise and emotion are very much intertwined, and some even view surprise as a basic emotion itself (Ekman et al., 1983). Others view surprise as an experience based on beliefs regarding the likelihood of an event (Lorini & Castelfranchi, 2007). Nevertheless, surprise might be a mixture of both, emotion and cognition, thereby bridging the two (Mellers et al., 2013). In general terms, surprise is experienced when people are confronted with stimuli that are not in accordance with their prior expectations (Noordewier et al., 2016). Various studies have demonstrated that the experience of surprise can be crucial for enhancing memory, as the discrepancy between what was expected and what occurs triggers learning and directs attention (Butterfield & Metcalfe, 2006; Rescorla & Wagner, 1972). For example, it was shown that surprising feedback increases attention which leads to better memory (Fazio & Marsh, 2009) and that memory can be predicted by the degree of expectancy violation (Greve et al., 2017). Also, developmental research has shown that children have better learning for objects and words presented after an expectancy violating event (Stahl & Feigenson, 2015, 2017). Nevertheless, not much is known about the difference between a positive and a negative surprise. A positive surprise is encountered when something goes against one's predictions but turns out better than expected. A negative

surprise is encountered when something goes against one's predictions but turns out worse than expected. Some studies suggest that positive and negative surprises differentially influence memory (De Loof et al., 2018; Jang et al., 2019), while other studies suggest that only the magnitude of a surprise is what influences memory (Fazio & Marsh, 2009; Rouhani et al., 2018).

Apart from surprise, emotional valence was also repeatedly shown to influence memory, as it is generally established that emotional events are remembered more vividly and accurately than neutral events (Tyng et al., 2017). However, it remains unclear to what extent positive or negative emotional content differentially influences learning and memory (Tyng et al., 2017). The current study investigates both surprise and emotional valence and aims to clarify their impact on episodic memory. For this purpose, a three-phase recognition experiment was carried out utilizing images with emotional valence. In phase 2, participants had to tell whether an image was old (from phase 1) or new. After every image the participants got feedback on whether their decision was correct or incorrect. This feedback was manipulated so that for half of the images participants received "correct" as their feedback and for the other half of the images "incorrect" as their feedback. Through this manipulation naturally positive and negative surprises occurred as even though participants might have been sure of their decision, the feedback they received was incorrect and vice versa.

Firstly, this study aims to investigate whether these positive or negative surprises differentially influence memory formation. It is hypothesized that images for which a positive surprise was encountered are generally better remembered than images for which a negative surprise was encountered. This hypothesis is based on classic theories of learning which predict that more attention is paid to stimuli with a positive outcome in order to better remember a possible reward in the future (Miendlarzewska et al., 2016; Schultz & Dickinson, 2000). Moreover, it is hypothesized that the magnitude of the surprise predicts the

memorability of an image, such that images for which a higher level of surprise was encountered are also remembered better. Furthermore, to extend the literature investigating the differential effect of positive versus negative stimuli on memory, the emotional valence of the images was taken into account. The current study also looks at the combination of emotional valence and the valence of surprise to explore possible interacting effects. This relation was, to our knowledge, not yet investigated. The current study's findings could be of great relevance in the educational setting and for learning in general, as giving the right feedback and using emotions is crucial for learning success.

Surprise and the Prediction Error

Even if we are sometimes not conscious of it, our brain continuously makes predictions based on our past experiences and memories to optimally utilize incoming information and select the most beneficial action (Friston, 2010). Making these predictions entails forming and updating internal models on the probability of events and occurrences in our environment, on which we base our decisions (O'Reilly et al., 2013). Naturally however, some events are predicted less well than others, which often leads to an experience of surprise for these unpredicted events (Noordewier et al., 2016). Thus surprise can be defined as the experience one encounters if an error in prediction is made and an event goes against prior expectations (Greve et al., 2017). Consequently, many studies operationalized surprise in terms of a prediction error (PE) (Fernández et al., 2016; Schultz, 2016).

A PE is the degree of conflict between a prediction and the actual encountered information (Greve et al., 2017; Schultz & Dickinson, 2000). The PE, and therefore surprise, is vital to human memory formation and updating. A PE signals a mismatch between our stored information and the actual information that occurs and thereby triggers learning in the brain (Fernández et al., 2016; Sinclair & Barense, 2018). This is in line with now-classic formal learning theories, which state that learning is proportional to the PE, or the difference

between expected and actual information and that learning occurs fastest when an event violates someone's expectations (Rescorla & Wagner, 1972). It is believed that the PE error has essential functions throughout the brain, being relevant for perception, approach, priming and several types of memory, including episodic memory (Henson & Gagnepain, 2010). Furthermore, PEs were found to be fundamentally important in domains such as reward learning (Pearce & Hall, 1980), decision making (Schultz & Dickinson, 2000) and associative memory formation (Greve et al., 2017). These widespread functions are thought to result from the PE because it modulates dopamine release in the brain (Schultz & Dickinson, 2000).

A distinction can be made between positive PEs and negative PEs, which according to Rescorla and Wagner (1972) and Schultz and Dickinson (2000), differentially influence dopamine release and learning. A positive PE is encountered when one makes a PE and gets positively surprised by the actual information. A negative PE is encountered when one makes a prediction and gets negatively surprised by the actual information (Fernández et al., 2016). Positive PEs, or in other words, when something turns out better than expected, are thought to increase the firing of dopaminergic neurons. In contrast negative PEs are thought to restrict the firing of dopaminergic neurons (Montague et al., 1996).

The modulation of dopamine release is especially relevant for learning and memory because increased dopamine levels were shown to promote synaptic plasticity in the hippocampus (Lemon & Manahan-Vaughan, 2006). The hippocampus is a structure located in the medial temporal lobe crucially responsible for encoding and retrieving episodic memories (Squire et al., 2004). An increase in synaptic plasticity, also known as long-term potentiation, is a process vital for the formation of long-term memories and can therefore explain how positive PEs enhance memory through dopamine modulation (Lemon & Manahan-Vaughan, 2006). From an evolutionary perspective, dopamine release for a positive PE act as a teaching signal. The predictive value of the preceding cue and the possibility to obtain a reward are

better remembered, thereby ultimately guaranteeing better evolutionary fitness. (Miendlarzewska et al., 2016).

In this regard, several studies could show that increasingly positive PEs as opposed to negative PEs indeed lead to increasingly better memory (De Loof et al., 2018; Jang et al., 2019). The theory that positive and negative PEs have a differential effect on learning and memory is commonly referred to as the signed effect of the PE, as the valence of the PE matters in this case. (Fernández et al., 2016). However, a different line of research has also found support for the theory of an unsigned effect of PEs on memory and learning (Fazio & Marsh, 2009; Rouhani & Niv, 2021; Rouhani et al., 2018). The unsigned effect of PEs implies that the magnitude of the PE, rather than its valence predicts enhanced memory for a surprising event. This means that positive and negative surprises are remembered equally well but that the absolute discrepancy between prediction and actual event influences memory and learning (De Loof et al., 2018).

Evidence for an unsigned effect of PE's on memory and learning stems from several accounts. Studies investigating the effect feedback has on subsequent memory have discovered and confirmed an effect which is often in the literature referred to as the hypercorrection effect (Butterfield & Metcalfe, 2001). The hypercorrection effect entails that when people are highly confident in their answer but make an error, they more readily and easily correct that error as long as they are given correct feedback (Butterfield & Metcalfe, 2006). Furthermore, high confidence answers that were corrected were also significantly better remembered, indicating that surprising feedback on the incorrectness of an answer can improve memory encoding (Butterfield & Metcalfe, 2006). By directly investigating the effects of surprising feedback on memory, Fazio and Marsh (2009) examined the mechanism underlying the hypercorrection effect and explored the difference between a positive and a negative surprise. In their experiment, participants answered general knowledge questions and were subsequently asked how confident they were in their answers. Participants received

feedback on their answers written in either green or red font for both, correct and incorrect feedback. In accordance with their predictions, Fazio and Marsh (2009) found that the color of the feedback and the content of the question was more often correctly remembered for high confidence errors and low confidence correct guesses. This finding is in line with the unsigned effect of the PE as both a positive surprise (low confidence correct guess) as well as a negative surprise (high confidence error) led to improved memory for the respective stimuli. Because participants also had better memory for the color of the feedback, an attribute they were not instructed to remember, Fazio and Marsh (2009) concluded that the hypercorrection effect and the effect of surprising feedback on memory is likely to occur due to increased attention being allocated to the surprising stimulus. This notion is in accordance with previous research that found that the experience of surprise interrupts ongoing thought processes and redirects attention upon the surprising stimulus to make sense of it (Horstmann, 2006).

Support for an unsigned effect of PEs on memory and learning also comes from a study by Rouhani et al. (2018). In their study, they specifically tested if signed or unsigned PEs differentially influence episodic memory. They presented participants with images of various scenes and let them, by trial and error, guess a predefined monetary reward each scene was associated with, whereby one category of scenes was generally associated with higher rewards and another category associated with lower rewards. As participants learned the associations between the scenes and got more confident in their answers, naturally positive and negative PEs occurred. In several experiments using this paradigm, Rouhani et al. (2018) could show that scenes, for which a large positive or negative PE was encountered, were remembered better in a subsequent recognition task. This demonstrates an unsigned effect of PEs and shows that the absolute magnitude of the PE predicts the influence of a PE on memory and learning. Contrary to the findings demonstrating an unsigned effect of PEs on memory, aforementioned studies on the neural mechanism underlying the link of PEs and dopamine release would predict a signed asymmetric effect, where negative PEs decrease

dopaminergic firing and positive PEs increase dopaminergic firing (Schultz & Dickinson, 2000).

However, advances in neuroscience have revealed that negative PEs or negative surprises lead to an activation of the locus coeruleus (LC), which releases norepinephrine to help allocate attentional resources and improve performance (Clewett et al., 2014). It was shown that these noradrenergic influences improve memory encoding for negative or unexpected feedback (Clewett et al., 2014). Furthermore, recent studies suggest that the LC not only releases norepinephrine but also releases dopamine in the hippocampus (Kempadoo et al., 2016). Some studies even suggest that the LC is the primary source of dopamine release in the hippocampus (Smith & Greene, 2012). This link between the noradrenergic system and dopamine release in the hippocampus provides a neural mechanism underlying findings demonstrating an unsigned effect of PEs on memory.

In a further study on the differential effect of signed and unsigned PEs on learning and memory, Rouhani and Niv (2021) could replicate their previous findings (Rouhani et al., 2018) and show that stimuli for which large unsigned PEs were encountered were better memorized throughout all experiments. Rouhani and Niv (2021) assume that the unsigned effect of PEs on memory is based on the heightened engagement of the LC, which releases norepinephrine and dopamine, thereby modulating hippocampal plasticity. However, in their experiments Rouhani and Niv (2021) could also show a signed effect of PEs on memory, that is, better memory for cues associated with a higher expected value.

As multiple separate findings show that both signed and unsigned PEs can enhance memory, it can be assumed that the effects occur in interaction through midbrain dopamine release and LC dopamine release, respectively. However, it is not yet evident how far unsigned and signed PEs differ in their effect on episodic memory. As previous studies focused on reward learning (Rouhani & Niv, 2021; Rouhani et al., 2018), associative learning (Greve et al., 2017) or memory for general knowledge questions (Fazio & Marsh, 2009), it

would be of great interest to explore how signed or unsigned PEs effect episodic memory in a classic image recognition task.

Another related phenomenon is that of incidental encoding of new "foil" items in a memory test. The first study using a memory-for-foils paradigm was conducted by Larry Jacoby and colleagues in 2005. In their study, Jacoby et al. (2005) presented participants with a random set of words, some of which should be processed on a deep semantic level (e.g. pleasantness of the word) and some on a more shallow, non-semantic, level (e.g. does the word contain e or u). In a second phase, an old/new recognition memory test was administered in which participants were tested on the deep semantic vs. shallow non-semantic words they studied earlier. In both memory tests (deep vs shallow) participants had to tell whether a word was part of the initial encoding or whether it was new (i.e. a foil words). In a final surprise recognition memory test, all semantic and non-semantic foil words of phase 2 were intermixed with completely new words and participants had to differentiate whether they already encountered a word or whether they have not seen it before. The results of the study showed that foil words which were part of the deep processing category were remembered better than foil words of the shallow processing category. This is remarkable because participants were not given processing instructions for any of the foil words but remembered words in the deep processing category better solely because they were processed in a more elaborate deep way. Thus, it was the cue presented in the phase 2 memory test (deep vs shallow) that initiated a deep vs shallow retrieval orientation and led to better incidental encoding of deep vs shallow foils (Jacoby et al., 2005). The foil effect could also play a role in the current study as it is also distributed in three phases. After participants have encoded images in phase 1, they had to tell in phase 2 whether an image was part of phase 1 or whether it is a new (foil) image. In phase 2 the current study induced positive and negative PEs of different strengths in participants by giving them feedback on their decision, similarly as Fazio and Marsh (2009). However, in contrast to Fazio and Marsh (2009), the current study

will also manipulate the feedback so that large positive or negative surprises occur whenever participants receive feedback that goes against their expectations or feedback that is surprisingly in line with their expectations. As a consequence, all foil images of phase 2 will be associated with a certain strengths and valence of surprise. Finally, in phase 3 all previous old and foil images are intermixed with completely new items to investigate whether the association of foil and old word with surprise has an influence on the memory for these images. Furthermore, this study investigates whether the magnitude or the sign of PEs is the primary factor for the influence of surprise on memory.

To our knowledge, no previous study investigated the differential effect of positive and negative PEs in an image recognition task, which is why the results of the current study can help to clarify the effect of positive and negative surprises on memory. This distinction of positive and negative surprise could have consequences for education and learning in general as giving the right feedback is crucial for learning success. Therefore, a general aim of the study is to extend the literature on signed versus unsigned PEs and yield unambiguous results on the effect of positive and negative surprise on memory. The hypotheses are:

H1: Unsigned effect: Stimuli for which a higher surprise in either positive or negative direction was encountered are remembered better than stimuli for which less surprise was encountered.

H2: Signed effect: Stimuli for which a positive surprise was encountered are remembered better than stimuli for which a negative surprise was encountered.

Emotional Valence

Next to surprise also emotion was frequently shown to influence memory and learning (Phelps, 2004; Um et al., 2012). Typically, emotion is classified within two continuous

dimensions, valence and arousal. Valence specifies how positive or negative an event is, arousal specifies how intense an event is (Lang et al., 1993). It was shown that emotional stimuli compared to neutral stimuli induce a “pop-out” like effect capturing attention, which has as a consequence that emotional stimuli are more likely to be encoded into long term memory (Vuilleumier, 2005; Yiend, 2010). Such an emotional memory effect could be confirmed by multiple studies which showed that emotional stimuli, like pictures, words or faces, are remembered better quantitatively as well as qualitatively than neutral stimuli (Buchanan & Adolphs, 2002; Kensinger & Corkin, 2003; Kensinger & Schacter, 2006).

The modulation of memories through emotions was formerly not as evident. It was assumed that the brain is organized in clearly separated neural systems where for instance the amygdala is responsible for emotional processes while the prefrontal cortex is responsible for cognition (Dolcos et al., 2011; Metcalfe & Mischel, 1999). However, recent research postulates that no area can be conceptualized as being specifically cognitive or affective as these neural systems are not reliant on a single brain region but are supported by a network of regions (Pessoa, 2008). It was shown that typical cognitive areas like the pre-frontal cortex are critically involved with emotion regulation and that emotions heavily influence cognition (Okon-Singer et al., 2015). Consequently it is not surprising that emotions influence memory systems and have a profound and long-term impact on memory formation and learning (Pessoa, 2008; Tyng et al., 2017).

That emotional stimuli are better remembered than neutral stimuli was, by most theories of emotional memory, attributed to arousal rather than to valence because positive and negative stimuli had similar effects on memory (Bowen et al., 2018; Dolcos et al., 2006). It is thought that emotional arousal improves memory because it captures attention and fosters elaboration of the stimuli (LaBar & Cabeza, 2006). Furthermore, emotional arousal was shown to be associated with greater activation of the amygdala, hippocampus and frontal as well as temporal areas (Murty et al., 2011) The strengthened connections of these areas,

which are crucial for memory formation, are thought to be the cause for the better memorability and persistence of emotionally arousing stimuli (Murty et al., 2011).

However, studies that match emotionally valent stimuli for arousal found evidence for an effect of valence which is independent of arousal (Bowen et al., 2018). Investigating the difference between positive and negative valence, Khairudin et al. (2011) and Khairudin et al. (2012) found that positive stimuli are being remembered better than negative stimuli with the authors concluding that negative valence may suppress explicit memory. Also Madan et al. (2019) showed that positive valence had an enhancing effect on memory whereas negative valence had an impairing effect. Furthermore it was shown that words associated with emotionally negative film clips were remembered less well than words associated with emotionally positive film clips (Anderson & Shimamura, 2005).

The aforementioned studies suggest that generally, positive stimuli are remembered better than negative stimuli. However, overall, the behavioral effects of valence on memory are mixed, such that multiple studies also found that negative rather than positive stimuli are remembered better. For instance, it was shown that participants remember negative stimuli more vividly than positive stimuli (Ochsner, 2000), and that they also recognize negative stimuli more often than positive stimuli (Kensinger et al., 2007). It was also shown that participants had better episodic memory of visual details of negative stimuli rather than positive stimuli. Furthermore, it was found that negative faces are better recognized than positive faces and that negative faces have better discriminability (Wang, 2013).

In their paper, Bowen et al. (2018) reviewed the differential effects of positive versus negative valence on memory. They found that overall, the amount of behavioral evidence for greater memorability and recognition of negative stimuli outweighs that of positive stimuli. A possible mechanism underlying this difference is that the encoding of negative stimuli, compared to positive stimuli, is more depended on sensory processes (Mickley & Kensinger, 2008). The enhanced utilization of sensory processes for stimuli of negative but not of

positive valence is in line with studies demonstrating that more attention is allocated towards negative stimuli (Simola et al., 2013) and explains why negative stimuli are remembered better quantitatively and qualitatively. Furthermore it was shown that valence also differentially influences neural processes, with only negative arousing stimuli but not positive arousing stimuli increasing connectivity within the amygdala (Kark & Kensinger, 2015). This is in line with a recent study that could detect differences in amygdala activation for memory retrieval of positively versus negatively valanced stimuli (Beyeler et al., 2016).

The previous studies demonstrate that emotional memory effects appear not only due to arousing effects of emotion but also due to the emotional valence. As especially negative valence was shown to affect memory processes the current study adheres to the model proposed by (Bowen et al., 2018) and predicts negative stimuli to be more frequently correctly recognized than positive stimuli. This study aims to extend the literature by demonstrating an effect of positive versus negative valence using the novel Open Affective Standardized Image Set (OASIS) that, to our knowledge, was not yet used in an picture recognition task investigating the effects of emotional valence (Kurdi et al., 2017). Previous studies investigating the effect valence has in an image recognition task have predominantly used the IAPS set which is considerably older and was already amply investigated on (for a review see Bowen et al. (2018)). Thus the current study aims at finding additional support for the results of (Bowen et al., 2018) while using a different and novel picture set. Confirming the theory with a different picture set can help to increase validity by showing that it is applicable to different contexts. The hypothesis regarding emotional valence is:

H3 Pictures of negative emotional valence will be more often correctly recognized than pictures of positive emotional valence.

Surprise and Emotional Valence

As mentioned above, some view surprise as a basic emotion (Ekman et al., 1983). However, unlike any other emotion, it is difficult to characterize surprise as being distinctly positive or negative in valence (Noordewier & Breugelmans, 2013). The initial affective reaction to surprise is driven by the unexpectedness of the stimulus and was shown to be of a mildly negative connotation (Noordewier & Breugelmans, 2013; Noordewier & van Dijk, 2019). This is because generally humans prefer consistency and predictability and why any surprise might at first be interpreted negatively (Bromberg-Martin & Hikosaka, 2009). However the actual affective reaction happens only a fraction later when the stimulus was made sense of and the valence of the surprising content leads to a clear positive or negative reaction (Noordewier et al., 2016). Both the valence of surprise and the valence of the stimulus have in common that they grab attention and foster elaboration (Noordewier et al., 2016; Vuilleumier, 2005). Thereby the surprise increases the more explanatory work has to be done in order to make sense of its outcomes (Foster & Keane, 2015, 2019). Previous studies investigating the effects of surprise on memory have mostly done so using neutral stimuli (Fazio & Marsh, 2009; Rouhani & Niv, 2021; Rouhani et al., 2018). However, as both emotional valence and positive and negative surprise were shown to differentially influence memory it is of great interest to see how both variables interact with each other. It can be assumed that being presented with stimuli of emotional valence influences the effect surprise has on memory, because especially negative stimuli were shown to be processed differentially than positive or neutral stimuli (Mickley & Kensinger, 2008; Simola et al., 2013). In this regard it is possible that being negatively surprised (high confidence error) for an emotionally negative stimulus has the most pronounced effect on memory. Consequently, if negative stimuli are remembered significantly better than positive stimuli and a high confidence error was made, the magnitude of the surprise should be higher than for positive stimuli or a

positive surprise. This is because more explanatory works has to be done to make sense of the mistake. Therefore, the hypothesis on the interaction of emotional valence and surprise is:

H4: Stimuli for which a negative surprise was encountered are better remembered if the stimuli were of negative emotional valence compared to positive emotional valence.

Method

Participants

A prior power analysis was conducted using G*Power (Faul et al., 2007) to estimate a sample size that yields sufficient statistical power for an effect size (d= .88) and alpha of .05. The effect size was based on the study of Fazio and Marsh (2009) who used a similar paradigm as they also investigated to what extent surprising feedback improves memory. The effect size was calculated with the help of an effect size calculator spreadsheet (Lakens, 2013). Results showed that a total sample of 19 participants is required to achieve a minimum power of .95. Based on the power analysis the overall sample consisted of 25 participants. Originally 28 participants successfully completed the study, but 3 participants were excluded because their accuracy in categorizing the images was significantly below chance level or because they had a strong (intentional) response bias. Of these 25 participants, 11 were male and 14 were female. The age of the participants ranged from 18 to 27 with a mean age of 21 years. Of the participants 10 were Dutch, 5 German, 1 French, 1 Belgian, 1 Ukrainian, 1 Spanish, 1 Hungarian, 1 Cypriot and 4 who preferred not to say. Of the participants, 11 stated that their highest education was school, 6 bachelor's degree, 3 a master's degree and 5 preferred not to say.

Participants were recruited by advertising the study directly in the Faculty of Social and Behavioral Sciences of Leiden University through the universities study participation system "SONA" (https://www.sona-systems.com) and by publishing invitations on social media

websites and via messenger apps. Only participants native or fluent in English with an age between 18-35 were recruited to ensure an even sample distribution. Furthermore, participants were excluded if they are using any (legal or illegal) psychoactive drugs or medication, have been diagnosed with any neurological or psychological disorders or are colorblind. Participants were asked to sign an informed consent and were compensated through credits for their participation. For the data collection and analysis, approval was requested of the ethical committee of Leiden University. Participants were debriefed about the purpose of the study after all data was collected.

Materials

For the purpose of the study an experiment was created using the program Psychopy (Peirce et al., 2019). The experiment was run on the online platform Pavlovia (pavlovia.org) and the overall study was conducted with Qualtrics (qualtrics.com). In total 320 images were used in the experiment. All pictures were taken from the Open Affective Standardized Image Set (OASIS) (Kurdi et al., 2017). The experiment consisted of three phases in which 80, 160 and 320 images were shown respectively. Phase two included all images of phase one together with the same number of new "foil" images. Phase three included all "old" and "foil" images of phase one and phase two and the same amount of completely new pictures. Before each phase participants read an instruction and performed a practice round consisting of five neutral pictures. Each phase consisted of the same amount of positively valanced and negatively valanced images. To achieve this the 160 most negatively valanced images and the 160 most positively valanced images were pre-selected and randomly assigned to each phase. Images with a very high positive or negative valence were excluded from the experiment in order to avoid pictures that stand out too much and to adhere to ethical guidelines. However, positive and negative images were matched for arousal such that positive and negative images

in every phase did not differ in their mean arousal level. For the three practice rounds before every phase 15 images with the most neutral valence were taken from the OASIS set.

Procedure

The overall procedure of the study can be divided into three phases. In phase 1, participants were presented with a randomized succession of 80 pictures of which 40 were of positive valence and 40 of negative valence. Each picture was shown to the participants for exactly two seconds. Between every picture a blank screen with a fixation cross was shown for one second. Participants were instructed that they would be presented with a series of pictures and that for each picture they had to indicate whether or not it contained a person. This simple task was chosen to ensure that participants engaged in the task.

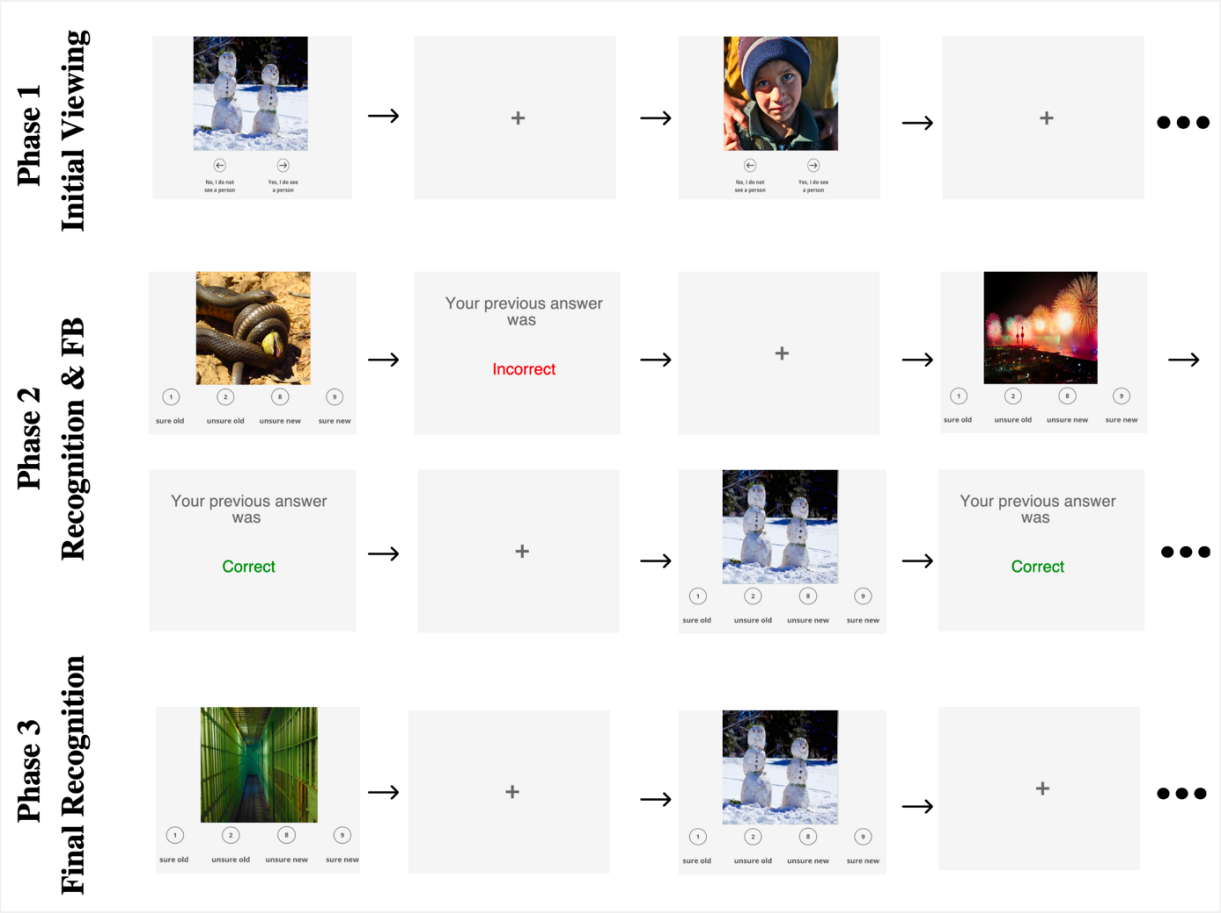
After phase 1, the participants were instructed for the recognition memory task in phase two. Participants were told that they will again see a succession of pictures in phase 2, of which some will be “old” previously seen pictures from phase 1 and soGTR13me will be “new” not previously seen pictures. Participants were instructed to indicate for each picture how confident they are that the picture is new or old on a 4-point scale. For this the participants were instructed the use the numbers 1, 2, 8 and 9 on their keyboard. The choices in the scale were “sure old”, “unsure old “, “unsure new” and “sure new”. The participants were also told that after every picture they will receive feedback on whether their decision was correct or incorrect. In phase two participants were shown 80 “new” pictures as well as the 80 ”old” pictures in a randomized order. Each picture is shown for two seconds with the scale of choices for the confidence rating being presented to them for three seconds below the picture. This means participants could view the picture for two seconds but had three seconds to give their response. Subsequently, participants saw feedback on the correctness of their decision for 1.5 seconds. The feedback said either “correct” or “incorrect” and was presented in green and red font respectively. However, the feedback was falsified, such that the

feedback presented was not dependent on the actual answer of the participant. This was done in order to evoke surprise in participants who might be confronted with incorrect feedback without being aware of it. Beforehand the feedback was randomly assigned to each image such that it is evenly distributed and that correct and incorrect feedback is given for the same amount of positive and negative pictures for each phase and within each category. For each participant the feedback on all images was newly randomized to avoid that each image is always paired with the same feedback. Additionally, the images within each phase of the experiment were randomized for every participant. Phase 2 resulted in four different possibilities regarding the feedback: A correct decision with unmanipulated feedback, an incorrect decision with unmanipulated feedback, a correct decision with manipulated feedback and an incorrect decision with manipulated feedback.

In phase 3, participants performed an old/new recognition memory task in which they were shown all previous 160 pictures from phase 2 and an additional 160 new pictures. Participants were instructed to once more differentiate between old pictures, which were all pictures from phase 2, and completely new pictures which they have never encountered before. In phase 3, participants viewed the pictures and the confidence choices for the same amount of time as in phase 2 but were able to switch to the next picture by giving their response. In phase 3, participants also did not receive any feedback on their decision. After phase 3, the participants were debriefed about the purpose of the study. The procedure of the study is illustrated in Figure 1.

Figure 1

Overview and Illustration of the different study phases



Note. In phase 1, a leftward arrow with the text "No, I don't see a person" and a rightward arrow with the text "Yes, I do see a person", was written below every image. Below every image in phase 2 the answer possibilities 1, 2, 8 and 9 were depicted with the captions "sure old", "unsure old", "unsure new" and "sure new", respectively. After every image in phase 2, the new screen with the feedback appeared. In phase 3, every image had below it the same answer possibilities as in phase 2.

Measures

The independent variable of emotional valence is specified to be either positive or negative depending on the respective picture. The variable surprise is specified as the discrepancy between the confidence judgement and the received feedback. It was differentiated between mild surprises and strong surprises and positive surprises and negative surprises. Thereby it was assumed that a strong positive surprise is encountered when the

confidence in the judgement was low but the feedback indicated the answer was correct. A mild positive surprise was assumed to be encountered when the confidence in the answer was high and the feedback was correct (no surprise). A strong negative surprise was assumed to be encountered if the confidence in a judgment is high but the feedback for the answer is incorrect. A mild negative surprise was assumed to be encountered if the confidence in the answer was low and the feedback was incorrect. Hence every old and foil image presented in phase 3 could be associated with one level of surprise encountered in phase 2. This was done so that in phase 3 it can be detected which level of surprise influences the correct recognition of the images. The left section of table 1 illustrates which responses in phase 2 were associated with which level of surprise depending on the content of the feedback. Furthermore, the right section of table 1 shows how the confidence rating of phase 3 was interpreted. For old and foil images ("old" was the correct answer for both in phase 3) the responses sure old, unsure old, unsure new and sure new were interpreted to be differentially correct. This means that the answer sure old was the most correct whereas the answer sure new was the least correct. Consequently, a new variable was created that does not only differentiate between a correct and an incorrect answer but also incorporates the confidence judgements of phase 3.

Table 1

Phase 2: response options and their associated level of surprise for old and foil images

Response Options	Feedback shown was "Correct"	Feedback shown was "Incorrect"
1 (Sure Old)	Mild positive surprise	Strong negative surprise
2 (Unsure Old)	Strong positive surprise	Mild negative surprise
8 (Unsure New)	Strong positive surprise	Mild negative surprise
9 (Sure New)	Mild positive surprise	Strong negative surprise

Results

In order to test hypothesis 1, which predicted that the magnitude of surprise decidedly influences memory (unsigned effect) and hypothesis 2 which predicted that the valence of surprise decidedly influences memory (signed effect), several repeated measures analyses were performed. The analyses were split into only foil images and only old images of phase 3. For all analyses, a repeated measures ANOVA was carried out, which included the four levels of surprise as specified in table 1. In this way, the repeated measures ANOVA included the two factors direction of surprise (positive or negative) and the strength of surprise (mild or strong).

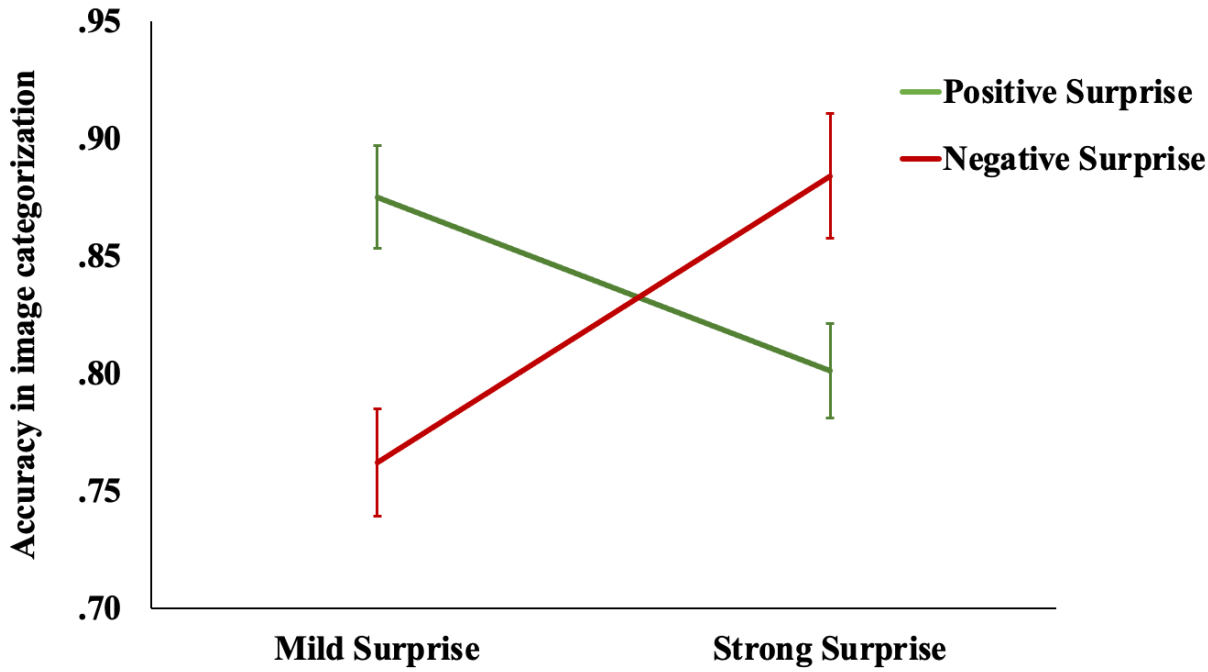
A repeated measures ANOVA on recognition memory for foil images in phase 3 showed no significant main effect of either the direction of surprise, $F(1, 24) = .16$ $p = .698$, $\eta 2 = .01$ or the strength of surprise, $F(1, 24) = .30$, $p = .586$, $\eta 2 = .01$. Also the interaction of the direction and the strength of surprise was not significant $F(1, 24) = 2.07$, $p = .164$, $\eta 2 = .08$. A repeated measures ANOVA on recognition memory for old items in phase 3 showed no significant main effect of either the direction of surprise, $F(1, 24) = .39$, $p = .539$, $\eta 2 = .09$, or the strength of surprise, $F(1, 24) = .94$, $p = .341$, $\eta 2 = .04$. However the interaction of the direction and the strength of surprise was significant $F(1, 24) = 11.81$, $p = .02$, $\eta 2 = .33$. Figure 2 illustrates this interaction for the correctness of only old images in phase 3 showing that when a mild surprise was encountered, mild positive surprises were remembered better than mild negative ones. Whereas when a strong surprise was encountered the strong negative surprise was remembered better than the strong positive one. Post-hoc paired sample t-tests (two-tailed) showed that the difference between mild positive and mild negative surprise was significant, $M = .11$ and SD = .21, $t(25) = 2.77$, $p = .011$, $d = .21$, and that the difference between strong positive and strong negative surprise was significant $M = -.08$ and SD = .17, $t(25) = -2.43$, $p < .023$, $d = .17$. This confirms the aforementioned relationship of positive

mild surprises being significantly better remembered than negative mild surprises and strong negative surprises being significantly better remembered than mild negative surprises.

However, since none of the four repeated measures analyses could find significant main effects of either the direction or the strength of surprise, the original hypotheses of a signed or an unsigned effect of surprise could not be confirmed. This means that neither direction nor the strength of a surprise could predict how well participants correctly remembered old or foil images in phase 3.

Figure 2

Interaction of the direction and the strength of surprise for only old images in phase 3



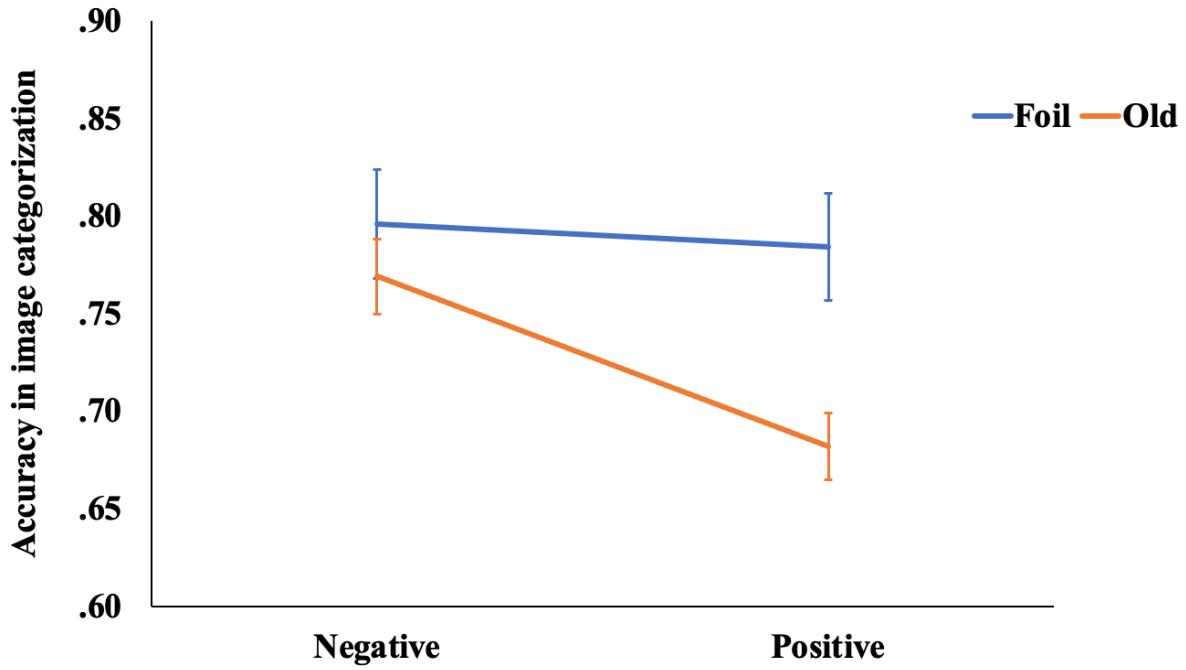
Note. The y-axis of the figure gives the mean of correctly categorized of images in phase 3. The x-axis gives the strength of surprise while the different lines give the direction of the surprise.

To investigate hypothesis 3, which predicted that images of negative valence are better remembered than images of positive valence, two repeated measures analyses were performed. This analysis included both the images from phase 2 as well as from phase 3. Again, a repeated measures ANOVA was conducted which included the factors emotional valence of an image (positive or negative), and image condition (old or foil). The repeated

measure analysis of images in phase 2 showed that the factor of emotional valence was significant, $F(1, 24) = 14.05$, $p < .001$, $\eta 2 = .37$ with negative images being more often correctly recognized than positive images. The condition of the image had no significant effect in phase 2, $F(1, 24) = 3.75$, $p = .0650$, $\eta 2 = .14$, but the interaction of emotional valence and image category was significant, $F(1, 24) = 4.75$, $p = 0.039$, $\eta 2 = .17$. Two (two-tailed) follow-up paired sample t-tests showed that the difference between negative foil and negative old images is not significant, $M = .02$ and SD = .17, $t(25) = .81$, $p = .426$, $d = .17$, while the difference between positive foil and positive old images is significant, $M = .11$ and SD = .21, $t(25) = 2.47$, $p = .021$, $d = .21$. This relationship is illustrated by Figure 3 which shows that old positive images are significantly less frequently correctly recognized than foil positive images.

Figure 3

The interaction of emotional valence and image category for images of phase 2.

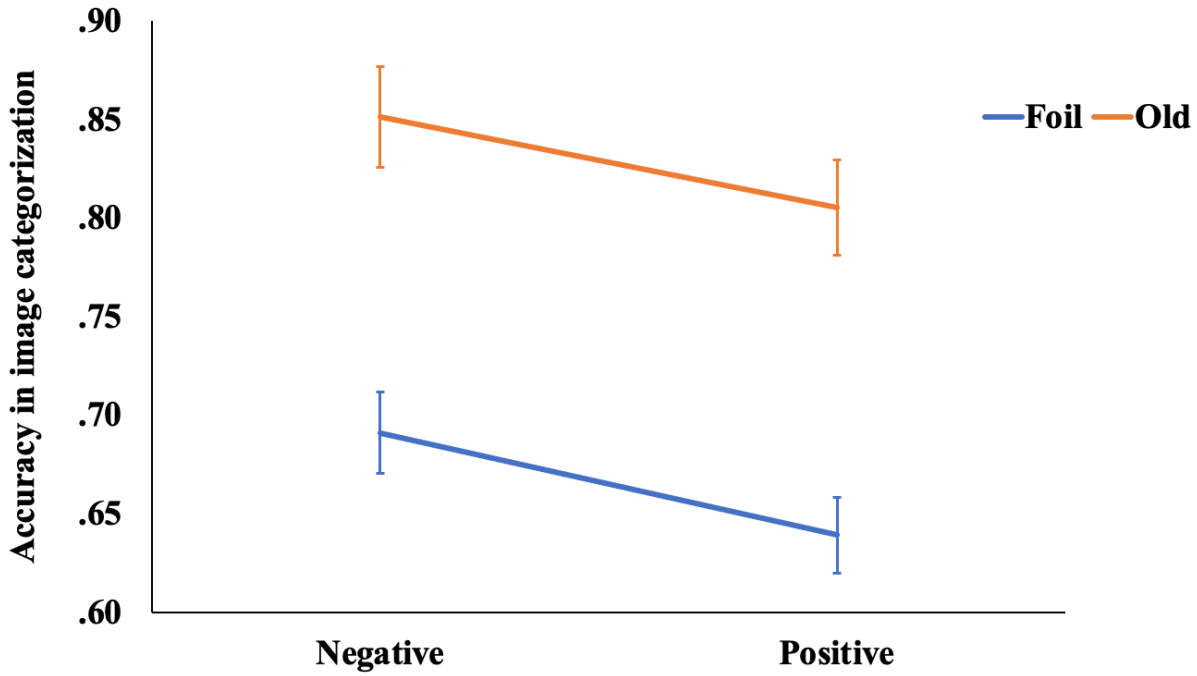


Note. The y-axis of the figure gives the mean of correctly categorized of images in phase 2. The x-axis gives the emotional valence while the different lines give the image condition.

The same repeated measures ANOVA performed for phase 2 images was also performed for images of phase 3. The analysis also showed that emotional valence predicts image recognition with negative images being significantly better recognized than positive images, $F(1, 24) = 13.24$, $p = .001$, $\eta 2 = .36$. These results as well as the results of phase 2 are in line with hypothesis 3 which predicted that negative images will be more frequently correctly recognized than positive images. In phase 3, contrary to phase 2, the image condition had a significant effect, $F(1, 24) = 112.19$, $p < .001$, $\eta 2 = .82$ with old images being significantly better correctly recognized than foil images. Furthermore, as Figure 4 illustrates, the interaction of emotional valence and image condition was shown to be not significant $F(1, 24) = .54$, $p = .818$, $\eta 2 = .01$.

Figure 4

The interaction of emotional valence and image condition for images of phase 3.



Note. The y-axis of the figure gives the mean of correctly categorized of images in phase 2. The x-axis gives the emotional valence while the different lines give the image condition.

In order to test hypothesis 4, which predicted that a negative surprise is better remembered for negative images than for positive images, two separate repeated measures ANOVA were performed. One looking at only images of positive emotional valence and one looking at only images of negative emotional valence for phase 3. Both analyses had two factors with two levels each: the direction of surprise (positive and negative) and the strength of surprise (mild and strong). However, for these analyses, trials for old and foil images were combined because if trials were split by the four levels of surprise, emotional valence and image category (old or foil), there would not be enough trials for all categories.

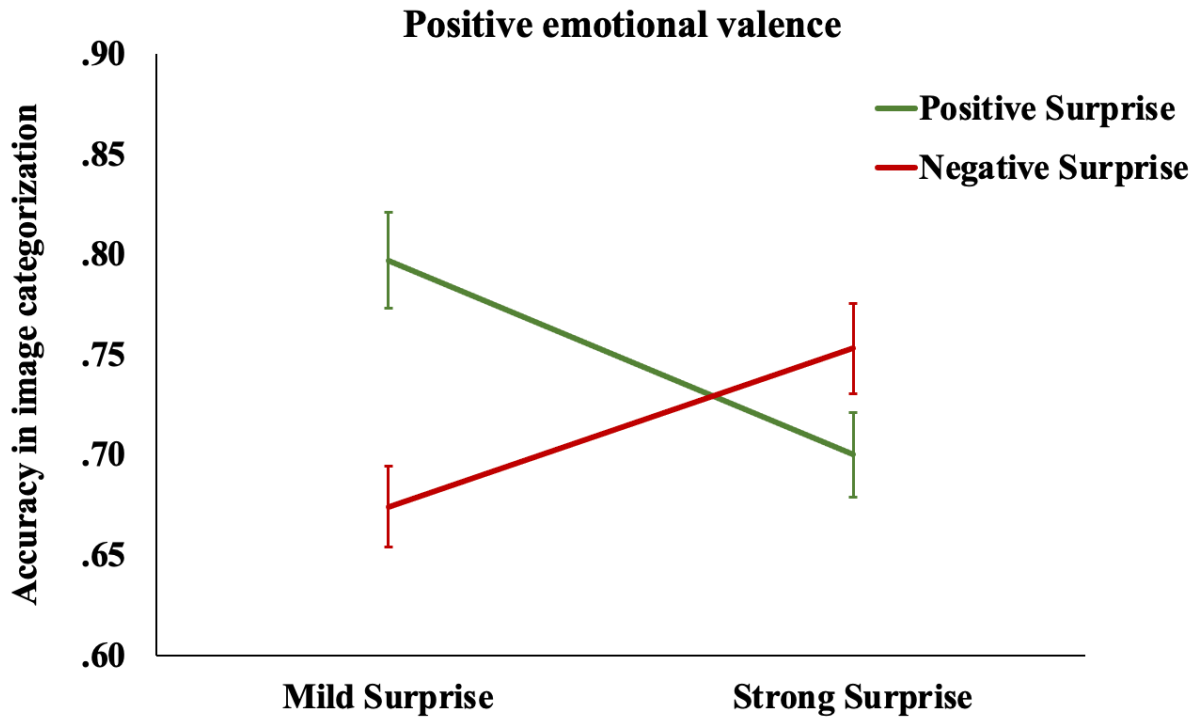
The analysis for only images of positive emotional valence resulted in no significant main effect of either the direction of surprise, $F(1, 24) = 1.83$, $p = .189$, $\eta 2 = .07$ or the strength of surprise, $F(1, 24) = .16$, $p = .690$, $\eta 2 = .01$. However the interaction between the direction of surprise and the strength of surprise was significant, $F(1, 24) = 15.52$, $p < .001$, $\eta 2 = .39$. Two post-hoc paired samples (two-tailed) t-tests confirmed the interaction, showing that for images of positive emotional valence a mild positive surprise was significantly better remembered than a mild negative surprise, $M = -.12$ and SD = .19, $t(25) = -3.25$, $p = .003$, $d = .19$, and a strong negative surprise was not significantly better remembered than a strong positive surprise, $M = .05$ and SD = .15, $t(25) = 1.80$, $p = .084$, $d = .15$.

For images with negative emotional valence, the two main effects of the direction of surprise $F(1, 24) = 0.67$, $p = .798$, $\eta 2 = .01$ and the strength of surprise $F(1, 24) = .94$, $p = .341$, $\eta 2 = .038$ were not significant. However, the interaction between the direction if surprise and the strength of surprise was significant, $F(1, 24) = 5.91$, $p = .023$, $\eta 2 = .19$. Two post-hoc paired samples (two-tailed) t-tests confirmed the interaction, showing that for images of negative emotional valence a mild positive surprise was not significantly better remembered than a mild negative surprise, $M = -.05$ and SD = .13, $t(25) = -1.79$, $p = .086$, $d = .13$, and a strong negative surprise was significantly better remembered than a strong positive surprise, $M = .09$ and SD = .20, $t(25) = 2.12$, $p = .084$, $d = .20$.

Figures 5 and 6 illustrate these significant interactions for images of negative and positive emotional valence respectively. It can be seen that for images of positive valence a mild positive surprise is significantly better remembered than a mild negative surprise while a strong negative surprise is not significantly better remembered than a mild negative surprise. For images of negative valence, the pattern reverses and a mild positive surprise is not significantly better remembered than a mild negative surprise, but a strong negative surprise is significantly better remembered than a strong positive surprise. These analyses could not confirm hypothesis 4 as the main effect of the direction of surprise had no significant effect when differentiating between positive and negative images. However, the significant interactions of the strength and the direction of surprise for both only positive and only negative images demonstrate that the emotional valence stands in interaction with the level of surprise in their effect on image recognition.

Figure 5

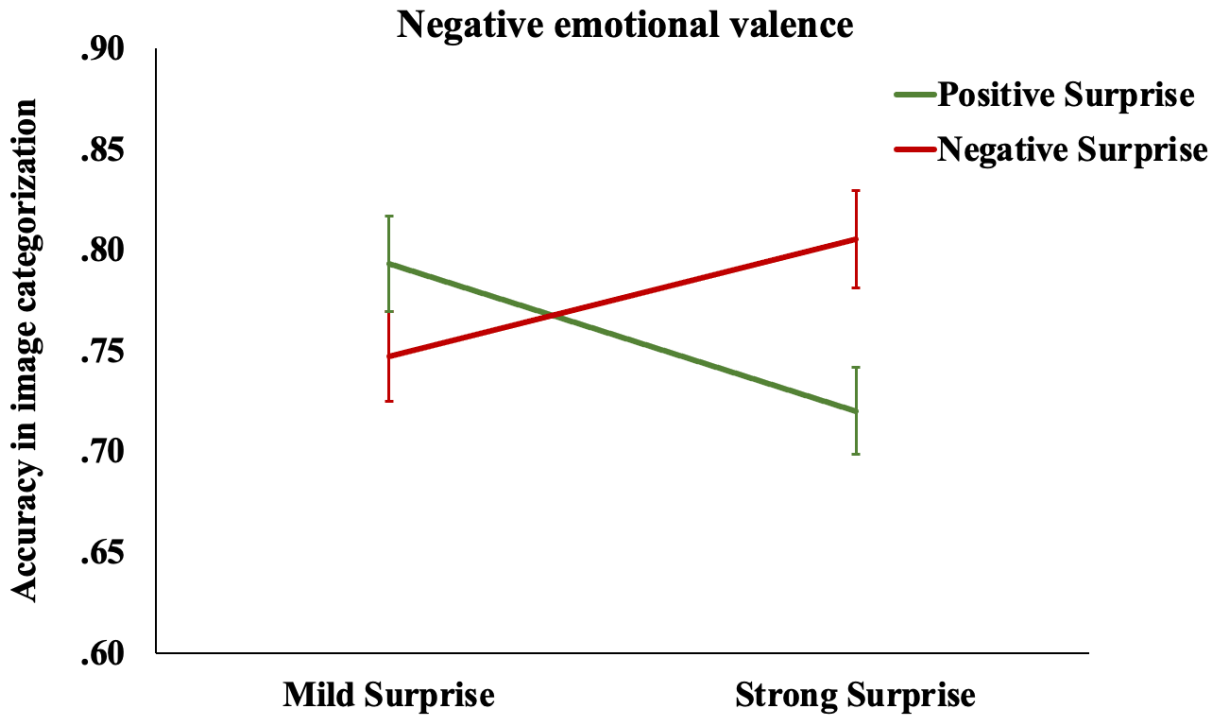
The interaction of direction and the strength of surprise for only positive images in phase 3



Note. The y-axis of the figure gives the mean of correctly categorized of images in phase 3. The x-axis gives the strength of surprise while the different lines give the direction of surprise.

Figure 6

The interaction of direction and the strength of surprise for only negative images in phase 3



Note. The y-axis of the figure gives the mean of correctly categorized of images in phase 3. The x-axis gives the strength of surprise while the different lines give the direction of surprise.

Discussion

Theoretical Implications

The aim of the current study was to investigate how different levels of surprise influence the memorability of images in a recognition memory task. To study this, we designed an experiment that manipulated the strength of surprise (mild vs strong) and the direction of surprise (positive vs negative). In addition, we were also interested in examining whether the emotional valence of the stimuli influenced subsequent recognition memory. Previous studies have demonstrated a significant unsigned effect of surprise on memory (Fazio & Marsh, 2009; Rouhani & Niv, 2021; Rouhani et al., 2018) or a significant signed effect of surprise on memory (De Loof et al., 2018; Jang et al., 2019). Studies demonstrating the unsigned effect of surprise on memory showed that the magnitude of a surprise is what

decidedly influences memory (Fazio & Marsh, 2009; Rouhani & Niv, 2021; Rouhani et al., 2018). This led to the prediction of the current study that strong positive and negative surprises will be better remembered than mild positive and negative surprises. Moreover by showing that positive surprises are generally better remembered than negative surprises, a different line of research has demonstrated a signed effect of surprise on memory which means that the direction of the surprise was shown to decidedly influence memory (De Loof et al., 2018; Jang et al., 2019). This led to the prediction of the current study that positive surprises will be generally better remembered than negative surprises. Our results showed that neither the strength of surprise nor the direction of surprise significantly influenced memory. Consequently, it could not be confirmed that stronger surprises are remembered better than milder surprises or that positive surprises are remembered better than negative surprises. Therefore, the results of this study are not in accordance with previous studies because no clear signed or unsigned effect of surprise on memory could be demonstrated.

Nevertheless, in the analyses looking at only old images a significant interaction of the direction and the strength of surprise was found. Images for which a mild positive surprise was encountered were significantly better remembered than images for which strong positive surprise was encountered and images for which a strong negative surprise was encountered were significantly better remembered than images for which a mild negative surprise was encountered.

To better make sense of this interaction, which is illustrated in Figure 2, it is important to understand that for the conception and the analysis of the study the four different levels of surprise, see Table 1, were regarded to be comparable with one another. This means, it was assumed that a mild positive surprise was of equal magnitude as a mild negative surprise and that a strong positive surprise was of equal magnitude as a strong negative surprise. However, the magnitude of surprise for a strong positive and a strong negative surprise can only hardly be equated because the formation of the two different levels of surprise is fundamentally

different: A strong negative surprise is encountered when you are very sure in your answer and receive “incorrect” as feedback whereas for a strong positive surprise you are unsure in your answer and receive “correct” as feedback. According to Foster & Keane (2015, 2019) the magnitude of a surprise can be best estimated by the amount of explanatory work one has to undertake in order to make sense of the surprising event. Therefore, the result that a strong negative surprise was significantly better remembered than a strong positive surprise is comprehensible. Subjectively seen, more explanatory work has to be carried out to explain being wrong when one was very sure in the answer than to explain being right when one was unsure of the answer. The magnitude of a strong negative surprise is even further increased when the feedback was manipulated, because more explanatory work has to be done to explain receiving “incorrect” as feedback when the answer was actually correct. The magnitude of a strong positive surprise on the other hand is not affected by manipulated feedback because participants were unsure in their answer and consequently did not expect a specific answer to be correct.

Furthermore, it is possible that compared to a strong negative surprise, a strong positive surprise might not pass a certain threshold in magnitude of surprise that would be necessary for an image to be better remembered. That a strong positive surprise is not comparable to and might not reach the same magnitude of surprise as a strong negative surprise might also be because of a negativity bias in their perception. The negativity bias describes the common effect that people generally learn from and attend more to negative rather than positive information (Baumeister et al., 2001; Soroka et al., 2019; Vaish et al., 2008). Consequently, a strong negative surprise might be perceived significantly different in magnitude than a strong positive surprise which makes them hard to compare and further clarifies why a strong negative surprise was significantly better remembered than a strong positive surprise.

Moreover, the results also showed that images associated with a mild positive surprise were significantly better remembered than images associated with a mild negative surprise. Similarly, as for both strong levels of surprise, both mild levels of surprise are not evenly comparable in the magnitude of their surprise. A mild negative surprise was assumed to be encountered when one is uncertain of an answer and receives "incorrect" as feedback. A mild positive surprise on the other hand was assumed to be encountered when one is certain of an answer and gets "correct" as feedback, which could actually be regarded as encountering no surprise at all. Consequently, their comparison is difficult and like the strong positive surprise they might not reach a high enough magnitude of surprise to significantly affect the memorization of the associated images. However, the question begs why a mild positive surprise, which can be subjectively as seen the least surprising, leads to a similarly high remembrance of associated images as a strong negative surprise? According to the negativity bias theory (Baumeister et al., 2001; Soroka et al., 2019; Vaish et al., 2008), a mild negative surprise should be better remembered than a mild positive surprise. A mild negative surprise is however not as negative as a strong negative surprise and hence the negativity bias could not be as effectual.

That images associated with a mild positive surprise were remembered well could be explained in terms of a confirmation bias (Nickerson, 1998; Tarantola, Folke, Boldt, Pérez, & de Martino, 2021; Vedejová & Čavojová, 2020). The confirmation bias states that information which is in accordance with our beliefs and expectations is favored, attended to more and also recalled better than information that is not in accordance with our beliefs (Nickerson, 1998; Tarantola et al., 2021; Vedejová & Čavojová, 2020). Furthermore, a recent study by Frost et al. (2015) showed that recognition memory is better for information that is in accordance with our beliefs compared to information that is not in accordance with our beliefs. Hence images for which a mild positive surprise was encountered could be remembered better because the participants' belief got confirmed, as one was sure in a decision and received "correct" as

feedback. This confirmation of one's beliefs could have led to an enhanced memorization of the respective images according to the confirmation bias.

Furthermore, this study also investigated whether images of positive or negative emotional valence are remembered better. Even though some previous research has shown that positive stimuli are remembered better than negative stimuli (Anderson & Shimamura, 2005; Khairudin et al., 2011, 2012; Madan et al., 2019), the majority of studies found that generally negative stimuli are remembered better than positive stimuli (Kensinger et al., 2007; Mickley Steinmetz, Knight, & Kensinger, 2016; Ochsner, 2000; Ritchey, Dolcos, & Cabeza, 2008; Sava, Paquet, Dumurgier, Hugon, & Chainay, 2016; Wang, 2013). Also, a metanalysis by Bowen et al. (2018) investigating whether positive or negative stimuli are generally remembered better came to the conclusion that evidence for negative stimuli being better remembered prevails. Therefore, hypothesis 3 predicted that the negative images in the study would be more frequently correctly recognized than the positive images. The results confirmed this hypothesis in showing that in both phase 2 and phase 3 of the experiment, images of negative emotional valence were significantly better remembered than images of positive emotional valence. Hence the current study is in line with and confirms findings of the metanalysis by Bowen et al. (2018) which demonstrated that in the majority of cases information of negative emotional valence is remembered and recognized better than information of positive emotional valence.

Additionally, this study set out to investigate whether there is a relationship between surprise and emotional valence in their effect on memory. It was predicted that images for which a negative surprise was encountered would be better remembered if they were of negative emotional valence compared to positive emotional valence. The results revealed that also when differentiating between images of positive and negative emotional valence the direction of surprise had no significant effect and therefore this hypothesis could not be confirmed. However, when looking at images of positive emotional valence and images of

negative emotional valence separately, an interaction between the direction of surprise and the strength of surprise was found for both positive and negative emotional valence (see Figure 5 and 6). For images of positive emotional valence there was a significant difference only for mild levels of surprise whereas for images of negative emotional valence there is a significant difference only for the strong levels of surprise. In other words, the data indicated that for images of positive emotional valence, a mild positive surprise is significantly better remembered than a mild negative surprise, whereas for images of negative emotional valence the opposite pattern was observed. It might be the case that not only the aforementioned confirmation bias but also the positive emotional valence drives the better remembrance for images of mild positive surprise over images of mild negative surprise. However, this interpretation should be treated with caution as it goes against the finding that negative images are generally better remembered than positive images and cannot be supported by literature. Furthermore, the analysis conducted to test hypothesis 4 showed that a strong negative surprise was significantly better remembered than a strong positive surprise for images of negative emotional valence but not for images of positive emotional valence. This might be the case because as images of negative emotional valence were shown to be remembered significantly better than images of positive emotional valence, the receipt of falsified incorrect feedback (feedback that the answer was incorrect when it was actually correct) might trigger a very strong negative surprise as the image was actually very well remembered. The better remembrance of an image of negative emotional valence itself in combination with a strong negative surprise might therefore explain why there is a significant difference in strong surprises for only images of negative emotional valence. This is in line with research postulating that a surprise is stronger and better remembered when more explanatory work has to be done in order to resolve it (Foster & Keane, 2015, 2019). So, even though the main effect of the direction of surprise when separating between emotional

valence was not significant this significant interaction gives some indication that hypothesis 4 might be correct.

Practical Implications

The term surprise, even though ubiquitous, might at first glance not be associated with much practical relevance. However, as surprise can also be defined as an error of prediction it is directly linked to learning. Even though the current study could not replicate an unsigned or a signed effect of surprise, it showed that when surprise was most likely encountered, in the case of a strong negative surprise, it significantly improves recognition memory. A finding that is of course not new, but once more demonstrates the well-established importance of immediate feedback for learning (Epstein et al., 2002). This is because feedback can trigger a surprise which in turn fosters active engagement and a deeper processing of the information by the learner, ultimately leading to better retention. The findings of this study could therefore be especially relevant in the educational setting where previous research on surprise has for instance demonstrated that disconfirmed predictions and being confused by a surprising contradiction can significantly boost learning (Brod et al., 2018; D'Mello et al., 2014). Moreover, this study also demonstrates the importance of positive feedback because being correct can boost learning through the confirmation bias. Consequently, the findings of the current study might be relevant for teachers as well as for students who try to optimize learning success. Furthermore, incorporating surprise as tool to remember something better could also benefit various other groups like the elderly or medical professionals. This means that knowing one did something correct can improve retention of something because it was in accordance with one's beliefs and therefore triggers the confirmation bias. Furthermore, this study also demonstrated bias towards negative information. The generally established negativity bias that led to negative images being more often correctly recognized than positive images in the current study, is a ubiquitous phenomenon which can greatly influence many

situations like witness accounts, assessments, and learning. Therefore, being conscious of this bias can be of great advantage.

Limitations and Future Research

As mentioned before, the conception of the study regarded the different levels of surprise to be comparable and of similar strength and direction. However, this assumption might have oversimplified the differences between surprise which ultimately also distorted the results. Even though positive and negative surprise are two directions of the same concept, their comparison is difficult as they entail vastly different premises and are of different strength. Therefore, it was perhaps too ambitious of the study to try and unify the different levels of surprise in one model. Future research should find a way in which surprise can be measured and specified more precisely. This would allow a better comparison of different levels of surprise and yield results that are more easily interpretable. The aim of future research should be to clearly differentiate between unsigned and signed effects and dissociate their impact on learning and memory. A further limitation of the study is the composition of the experiment. It was assumed that participants were surprised because of the feedback on their decision in phase 2. However, the number of trials (160) and the limited time of exposure to the feedback (1.5 seconds) could have reduced the feedbacks surprising effect as participants already awaited the next image. Furthermore, an additional question at the end of the study which asked participants whether they have noticed that the feedback was at times falsified revealed that 75% of all participants thought the feedback was manipulated. It might be that some participants ignored the feedback after they noticed it was fake and therefore also did not get surprised. Future studies investigating the effect of surprising feedback should therefore either avoid manipulated feedback all together or have more trials and await natural occurrences of, for instance, high confidence incorrect answers.

Conclusion

Conversely to the literature, the study could not demonstrate a clear unsigned or signed effect of surprise on memory. However, it was shown that a strong negative surprise led to significantly enhanced memory for associated images. Interestingly no surprise at all was also shown to significantly enhance memory for associated images which might be due to a confirmation bias. Furthermore, this study supported existing literature showing that images of negative emotional valence are generally better remembered than images of positive emotional valence. Moreover, the current study investigated the interaction of surprise and emotional valence in their effect on memory. No significant relation was found, but the results indicate the emotional valence of an image has an influence on the effect a surprise has on memorability. For instance, a strong negative surprise was only significantly better remembered than a strong positive surprise for negative but not positive images. Overall, this study yielded interesting results in showing that only a strong negative surprise enhances memory for associated images but that also no surprise at all led to better memory a finding that can be of great relevance in education and learning research.

Acknowledgements

This study was supported by the faculty of Social and Behavioral Sciences of Leiden University and supervised by Dr. David Amadeus Vogelsang. The design of the online experiment and the data collection was carried out in cooperation with Ilona Enyedi.

References

Anderson, L., & Shimamura, A. P. (2005). Influences of emotion on context memory while viewing film clips. *American Journal of Psychology*.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*. https://doi.org/10.1037//1089-2680.5.4.323

Beyeler, A., Namburi, P., Glober, G. F., Simonnet, C., Calhoon, G. G., Conyers, G. F., … Tye, K. M. (2016). Divergent Routing of Positive and Negative Information from the Amygdala during Memory Retrieval. *Neuron*. https://doi.org/10.1016/j.neuron.2016.03.004

Bowen, H. J., Kark, S. M., & Kensinger, E. A. (2018). NEVER forget: negative emotional valence enhances recapitulation. *Psychonomic Bulletin and Review*, *25*(3), 870–891. https://doi.org/10.3758/s13423-017-1313-9

Brod, G., Hasselhorn, M., & Bunge, S. A. (2018). When generating a prediction boosts learning: The element of surprise. *Learning and Instruction*, *55*(January), 22–31. https://doi.org/10.1016/j.learninstruc.2018.01.013

Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*. https://doi.org/10.1016/j.neuron.2009.06.009

Buchanan, T. W., & Adolphs, R. (2002). The role of the human amygdala in emotional modulation of long-term declarative memory. https://doi.org/10.1075/aicr.44.02buc

Butterfield, B., & Metcalfe, J. (2001). Errors Committed with High Confidence Are Hypercorrected. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/0278-7393.27.6.1491

Butterfield, B., & Metcalfe, J. (2006). The correction of errors committed with high confidence. *Metacognition and Learning*, *1*(1), 69–84. https://doi.org/10.1007/s11409-006-6894-z

Clewett, D., Schoeke, A., & Mather, M. (2014). Locus coeruleus neuromodulation of memories encoded during negative or unexpected action outcomes. *Neurobiology of Learning and Memory*, *111*, 65–70. https://doi.org/10.1016/j.nlm.2014.03.006

Conway, M. A., Anderson, S. J., Larsen, S. F., Donnelly, C. M., McDaniel, M. A., McClelland, A. G. R., … Logie, R. H. (1994). The formation of flashbulb memories. *Memory & Cognition*, *22*(3), 326–343. https://doi.org/10.3758/BF03200860

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170. https://doi.org/10.1016/j.learninstruc.2012.05.003

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*. https://doi.org/10.1371/journal.pone.0189212

Dolcos, F. (2006). Neural correlates of emotional evaluation and emotional episodic memory: Electrophysiological and hemodynamic evidence. *Dissertation Abstracts International: Section B: The Sciences and Engineering*.

Dolcos, Florin, Iordan, A. D., & Dolcos, S. (2011). Neural correlates of emotion - Cognition interactions: A review of evidence from brain imaging investigations. *Journal of Cognitive Psychology*, *23*(6), 669–694. https://doi.org/10.1080/20445911.2011.594433

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science*. https://doi.org/10.1126/science.6612338

Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *Psychological Record*. https://doi.org/10.1007/BF03395423

Faul, F., ErdFelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3.1 manual. *Behavioral Research Methods*.

Fazio, L. K., & Marsh, E. J. (2009). Surprising feedback improves later memory. *Psychonomic Bulletin and Review*, *16*(1), 88–92. https://doi.org/10.3758/PBR.16.1.88

Fernández, R. S., Boccia, M. M., & Pedreira, M. E. (2016). The fate of memory: Reconsolidation and the case of Prediction Error. *Neuroscience and Biobehavioral Reviews*, *68*, 423–441. https://doi.org/10.1016/j.neubiorev.2016.06.004

Foster, M. I., & Keane, M. T. (2015). Why some surprises are more surprising than others: Surprise as a metacognitive sense of explanatory difficulty. *Cognitive Psychology*, *81*, 74–116. https://doi.org/10.1016/j.cogpsych.2015.08.004

Foster, M. I., & 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

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Frost, P., Casey, B., Griffin, K., Raymundo, L., Farrell, C., & Carrigan, R. (2015). The Influence of Confirmation Bias on Memory and Source Monitoring. *Journal of General Psychology*, *142*(4), 238–252. https://doi.org/10.1080/00221309.2015.1084987

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

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315–1326. https://doi.org/10.1002/hipo.20857

Horstmann, G. (2006). Latency and duration of the action interruption in surprise. *Cognition and Emotion*. https://doi.org/10.1080/02699930500262878

Jacoby, L. L., Shimizu, Y., Velanova, K., & Rhodes, M. G. (2005). Age differences in depth of retrieval: Memory for foils. *Journal of Memory and Language*. https://doi.org/10.1016/j.jml.2005.01.007

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

Kark, S. M., & Kensinger, E. A. (2015). Effect of emotional valence on retrieval-related recapitulation of encoding activity in the ventral visual stream. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2015.10.014

Kempadoo, K. A., Mosharov, E. V., Choi, S. J., Sulzer, D., & Kandel, E. R. (2016). Dopamine release from the locus coeruleus to the dorsal hippocampus promotes spatial learning and memory. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(51), 14835–14840. https://doi.org/10.1073/pnas.1616515114

Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: Are emotional words more vividly remembered than neutral words? *Memory and Cognition*. https://doi.org/10.3758/BF03195800

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2007). Effects of emotion on memory specificity in young and older adults. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*. https://doi.org/10.1093/geronb/62.4.P208

Kensinger, E. A., & Schacter, D. L. (2006). Processing emotional pictures and words: Effects of valence and arousal. *Cognitive, Affective and Behavioral Neuroscience*, *6*(2), 110–126. https://doi.org/10.3758/CABN.6.2.110

Khairudin, R., Givi, M. V., Wan Shahrazad, W. S., Nasir, R., & Halim, F. W. (2011). Effects of emotional contents on explicit memory process. In *Pertanika Journal of Social Science and Humanities*.

Khairudin, R., Nasir, R., Halim, F. W., Zainah, A. Z., Wan Shahrazad, W. S., Ismail, K., & Valipour, G. M. (2012). Emotion and explicit verbal memory: Evidence using Malay Lexicon. *Asian Social Science*. https://doi.org/10.5539/ass.v8n9p38

Kraha, A., & Boals, A. (2014). Why so negative? Positive flashbulb memories for a personal

event. *Memory*, *22*(4), 442–449. https://doi.org/10.1080/09658211.2013.798121

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the Open Affective Standardized Image Set (OASIS). *Behavior Research Methods*. https://doi.org/10.3758/s13428-016-0715-3

LaBar, K. S., & Cabeza, R. (2006). Cognitive neuroscience of emotional memory. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn1825

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2013.00863

Lang, P. J., Greenwald, M. K., & Bradley, M. M. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology - Wiley Online Library.*

Lemon, N., & Manahan-Vaughan, D. (2006). Dopamine D1/D5 receptors gate the acquisition of novel information through hippocampal long-term potentiation and long-term depression. *Journal of Neuroscience*, *26*(29), 7723–7729. https://doi.org/10.1523/JNEUROSCI.1454-06.2006

Lorini, E., & Castelfranchi, C. (2007). The cognitive structure of surprise: Looking for basic principles. *Topoi*. https://doi.org/10.1007/s11245-006-9000-x

Madan, C. R., Scott, S. M. E., & Kensinger, E. A. (2019). Positive emotion enhances association-memory. *Emotion*, *19*(4), 733–740. https://doi.org/10.1037/emo0000465

Mellers, B., Fincher, K., Drummond, C., & Bigony, M. (2013). *Surprise. A belief or an emotion? Progress in Brain Research* (1st ed., Vol. 202). Elsevier B.V. https://doi.org/10.1016/B978-0-444-62604-2.00001-0

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review*. https://doi.org/10.1037/0033-295X.106.1.3

Mickley, K. R., & Kensinger, E. A. (2008). Emotional valence influences the neural correlates

associated with remembering and knowing. *Cognitive, Affective and Behavioral Neuroscience*. https://doi.org/10.3758/CABN.8.2.143

Mickley Steinmetz, K. R., Knight, A. G., & Kensinger, E. A. (2016). Neutral details associated with emotional events are encoded: evidence from a cued recall paradigm. *Cognition and Emotion*. https://doi.org/10.1080/02699931.2015.1059317

Miendlarzewska, E. A., Bavelier, D., & Schwartz, S. (2016). Influence of reward motivation on human declarative memory. *Neuroscience and Biobehavioral Reviews*, *61*, 156–176. https://doi.org/10.1016/j.neubiorev.2015.11.015

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–1947. https://doi.org/10.1523/jneurosci.16-05-01936.1996

Murty, V. P., Ritchey, M., Adcock, R. A., & LaBar, K. S. (2011). Reprint of: fMRI studies of successful emotional memory encoding: A quantitative meta-analysis. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2011.02.031

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*. https://doi.org/10.1037/1089-2680.2.2.175

Noordewier, M. K., & Breugelmans, S. M. (2013). On the valence of surprise. *Cognition and Emotion*, *27*(7), 1326–1334. https://doi.org/10.1080/02699931.2013.777660

Noordewier, M. K., Topolinski, S., & Van Dijk, E. (2016). The Temporal Dynamics of Surprise. *Social and Personality Psychology Compass*, *10*(3), 136–149. https://doi.org/10.1111/spc3.12242

Noordewier, M. K., & van Dijk, E. (2019). Surprise: unfolding of facial expressions. *Cognition and Emotion*, *33*(5), 915–930. https://doi.org/10.1080/02699931.2018.1517730

O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior

cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38). https://doi.org/10.1073/pnas.1305373110

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General*. https://doi.org/10.1037//0096-3445.129.2.242

Okon-Singer, H., Hendler, T., Pessoa, L., & Shackman, A. J. (2015). The neurobiology of emotion-cognition interactions: Fundamental questions and strategies for future research. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2015.00058

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*. https://doi.org/10.1037/0033-295X.87.6.532

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-01193-y

Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*. https://doi.org/10.1038/nrn2317

Phelps, E. A. (2004). Human emotion and memory: Interactions of the amygdala and hippocampal complex. *Current Opinion in Neurobiology*. https://doi.org/10.1016/j.conb.2004.03.015

Rescorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement BT - Clasical conditioning II: current research and theory. *Clasical Conditioning II: Current Research and Theory*.

Ritchey, M., Dolcos, F., & Cabeza, R. (2008). Role of amygdala connectivity in the persistence of emotional memories over time: An event-related fMRI investigation. *Cerebral Cortex*. https://doi.org/10.1093/cercor/bhm262

Rouhani, N., & Niv, Y. (2021). Signed and unsigned reward prediction errors dynamically

enhance learning and memory. *ELife*, *10*(Lc), 1–28. https://doi.org/10.7554/eLife.61077

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

Sava, A. A., Paquet, C., Dumurgier, J., Hugon, J., & Chainay, H. (2016). The role of attention in emotional memory enhancement in pathological and healthy aging. *Journal of Clinical and Experimental Neuropsychology*. https://doi.org/10.1080/13803395.2015.1123225

Schultz, W. (2016). Dopamine reward prediction-error signalling: A two-component response. *Nature Reviews Neuroscience*, *17*(3), 183–195. https://doi.org/10.1038/nrn.2015.26

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience*. https://doi.org/10.1146/annurev.neuro.23.1.473

Simola, J., Torniainen, J., Moisala, M., Kivikangas, M., & Krause, C. M. (2013). Eye movement related brain responses to emotional scenes during free viewing. *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/fnsys.2013.00041

Sinclair, A. H., & Barense, M. D. (2018). Surprise and destabilize: Prediction error influences episodic memory reconsolidation. *Learning and Memory*, *25*(8), 369–381. https://doi.org/10.1101/lm.046912.117

Smith, C. C., & Greene, R. W. (2012). CNS dopamine transmission mediated by noradrenergic innervation. *Journal of Neuroscience*, *32*(18), 6072–6080. https://doi.org/10.1523/JNEUROSCI.6486-11.2012

Soroka, S., Fournier, P., & Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1908369116

Squire, L. R., Stark, C. E. L., & Clark, R. E. (2004). The medial temporal lobe. *Annual*

Review of Neuroscience. https://doi.org/10.1146/annurev.neuro.27.070203.144130

Stahl, A. E., & Feigenson, L. (2015). Observing the unexpected enhances infants' learning and exploration. *Science*. https://doi.org/10.1126/science.aaa3799

Stahl, A. E., & Feigenson, L. (2017). Expectancy violations promote learning in young children. *Cognition*. https://doi.org/10.1016/j.cognition.2017.02.008

Tarantola, T., Folke, T., Boldt, A., Pérez, O. D., & de Martino, B. (2021). Confirmation bias optimizes reward learning. *BioRxiv*. https://doi.org/10.1101/2021.02.27.433214

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, *8*(AUG). https://doi.org/10.3389/fpsyg.2017.01454

Um, E. R., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional Design in Multimedia Learning. *Journal of Educational Psychology*. https://doi.org/10.1037/a0026609

Vaish, A., Woodwaard, A., & Grossmann, T. (2008). Not All Emotions Are Created Equal: The Negativity Bias in Social-Emotional Development. *Psychological Bulletin*, *134*(3), 383–403. https://doi.org/10.1037/0033-2909.134.3.383

Vedejová, D., & Čavojová, V. (2020). How to examine confirmation bias? *Ceskoslovenska Psychologie*.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2005.10.011

Wang, B. (2013). Facial expression influences recognition memory for faces: Robust enhancement effect of fearful expression. *Memory*. https://doi.org/10.1080/09658211.2012.725740

Yiend, J. (2010). The effects of emotion on attention: A review of attentional processing of emotional information. *Cognition and Emotion*. https://doi.org/10.1080/02699930903205698