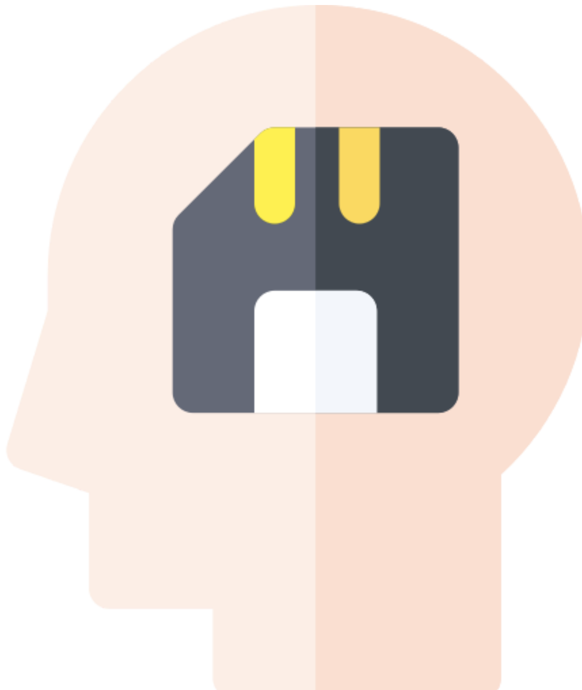
Emotional MetaMemory



How emotional valence and arousal influence metacognition for memory

Author: Sebastian Scott Engen (Student number: 201708490)

Supervisor: Micah Allen

Bachelor project in Cognitive Science

School of Communication and Culture, University of Aarhus

20. December 2019

Characters: 43378 (18.5 pages including figures)

Summary

Emotion and cognition have been found to interact during decision making and this study will analyse how emotions systematically manipulate a certain form of cognition, namely metamemory; the concept of how people monitor, assess, manage and relate to their own memory capabilities and strategies (“Koriat: Metamemory: The feeling of knowing and its vagaries—Google Scholar,” n.d.; Nelson, 1990; Pannu & Kaszniak, 2005)

Albeit metamemory has been studied in many research laboratories, few studies explicitly investigate how metamemory works on emotionally charged or valenced information(Fairfield, Mammarella, Palumbo, & Di Domenico, 2015). Moreover, the research that has been carried out have also been limited to statements of contradictory findings; where different dimensions of emotions either all rendered a heightened effect on recall and metamemory, interacted or showed no effects of emotion at all(Allen et al., 2016; Fairfield et al., 2015; Massoni, 2014; Sharot, 2011).

It is therefore hard to draw any clear conclusions and that is why this thesis sets out to investigate more clearly the effects of emotions, in the dimensions ‘Valence’ and ‘Arousal’, to get at the underlying generalizable systematicities of the influences over memory.

This study includes a meta-memory experiment that had participants memorize 50 words followed by a cued-recall task and confidence rating. The stimuli was 1200 English words selected from the “Affective Norms for English Words” (ANEW) database (Bradley, Lang, Bradley, & Lang, 1999) employed in a 2x2 design on the emotional dimensions of Arousal {Low,High} and Valence {Negative,Positive}.

The experiment assessed metacognition in terms of signal detection theoretic measures of sensitivity, metacognitive sensitivity, and metacognitive efficiency, leveraging the hierarchical bayesian model of Stephen Fleming (Fleming, 2017,).

The Results of the primary analysis show a strong positive effect of Valence on both memory performance and metacognition supporting a role for emotions in guiding confidence and memory performance. The results of the physiological monitoring moreover suggests that the heart hold a mediatory role of these valence effects.

However, this study found little evidence for an independent effect of arousal on discrimination and metacognition, contrary to prior studies (Allen et al., 2016; de Gee et al., 2017;Delfini & Campos, 1972). However, effects were found for Arousal interacting with Valence.

Future work will be to scale up the experiment to explore the effects of Valence further. It is also suggested that future work adopts the full hierarchical model of (Fleming & Lau, 2014) to render the metacognitive efficiency interpretable (mratio). Ultimately, a full-blown replication of the ANEW database in Danish would also be wise moving forward.

Summing up, this study sheds light on the impact of valence and arousal on learning new words. It shows how emotional words impact what subjects remember and how well they remember it.

Keywords: Emotion, arousal, valence, heart rate variability, metacognition, metamemory, memory

Introduction

Theoretical

People synthesize their emotions and cognition into a coherent perception of reality, but how this happens is a question with a considerable research effort, taking place in a maturing field of Cognitive Science (Fig.2). The study of how emotion and cognition interact has a long history in western thought: the first endeavors to investigate the interplay between the rational and the emotional was first led by philosophers (Plato and The Stoics)(Scherer, 2011), who established a dissociation between the emotional animal part of the individual and its rational human counterpart. This antagonism was further taken up in early and late christianity, and early modern philosophy (Kant and Descartes) (Stanford Encyclopedia of Philosophy). where 'romantic infatuation' for instance was stamped as an illness of the mind and detachment from one's emotions, or so called "agitations" and "disturbances" were stamped as a higher good.

Today, however, contemporary cognitive science has taken a turn towards a more integrated approach where affective aspects of psychological function are crucial and useful aspects of cognition and adaptive decision-making ("Damasio: Descartes' Error; Ledoux, 1998; Seo, Goldfarb, & Barrett, 2010). Publications now question classic studies such as the work on Prospect theory carried out by Kahneman and Twersky, who failed to incorporate how affective states systematically manipulate the dimensions of gains and losses (Campos-Vazquez & Cuilty, 2014). Other publications clue to how cognitive and moral enhancement can be achieved by leveraging people's emotional lives (Huebner, Dwyer, & Hauser, 2009) and some studies even show how our relation to time itself is adapting to emotional fluctuations (Liu, Feng, Chen, & Li, 2013). The landscape of emotional research is definitely changing and "Scientists now assume that emotions are, for better or worse, the dominant driver of most meaningful decisions in life." (Lerner, Li, Valdesolo, & Kassam, 2015).

This study examines the interplay of emotion and cognition within the topic of metacognition: how people relate to their own cognition through introspection (Metacognition: Knowing about Knowing). The word 'Meta' stems from the greek 'beyond' and hints at something that abstracts away from the original. To help grasp the idea one can look to the 'Tip Of The Tongue' (TOT) phenomenon: it refers to the experience of confidence that one has the answer to something, like the date of one's grandmother's birthday, yet is unable to produce the actual answer, as it lingers on the cusp of conscious awareness (Brown & McNeill, 1966). Metacognition can be viewed as the conceptual encapsulating of this phenomena of relating to one's own knowledge.

However, more specifically, we'll concern ourselves with metamemory in this study. Metamemory is the form of metacognition that concerns itself with how people monitor, assess, manage and relate to their own memory capabilities and strategies ("Koriat: Metamemory: The feeling of knowing; Nelson, 1990; Pannu & Kaszniak, 2005). The investigation of how people relate to their memories stretch all the way back to ancient Rome where people investigated their memories like they were artifacts displayed in museums ("Method of loci—Wikiwand," n.d.). Nevertheless, in the 1970's a formalised scientific investigation in this field took hold with research on children and their cognitive development led by John H. Flavel, who also wrote a brief treatise on metacognition framing it as a field “worthy of study” (Flavell, 1979; Flavell & Wellman, 1975). The interest in this topic soon grew as curious and paradoxical findings accumulated suggesting that people were strangely accurate in their predictions of how well they’d remember or forget studied information, which drove the development of new methodologies to empirically investigate people's metamemory (Nelson, 1990). Contemporary research, springing off from these foundational works emphasizing a degree of control over/insight into the memory-machinery, have now started to investigate more actively, how the construction and the quality of memory can be altered by emotional components. This is a research effort is motivated under the central assumption that one’s memories and one’s perception of them guides one's future behaviour, and thereby the promise of how better knowledge on this topic will be able to help aid and tune humans wherever memory is crucially leveraged, as it can hold huge implications for people's health, education, and worklife (Dunning, Heath, & Suls, 2016; “Metamemory—Psychology—Oxford” n.d.).

Diving into the landscape of emotion research, one finds a bedrock of common understanding among theorists that emotional states can be arranged along two bipolar axises of ‘Arousal’ and ‘Valence’ (Barrett & Russell, 1999)(Fig.1). Arousal refers to the intensity of an affective state, grounded in measurable changes in parasympathetic nervous system activation, where Valence reflects the positive-to-negative subjective evaluation of an experienced state.

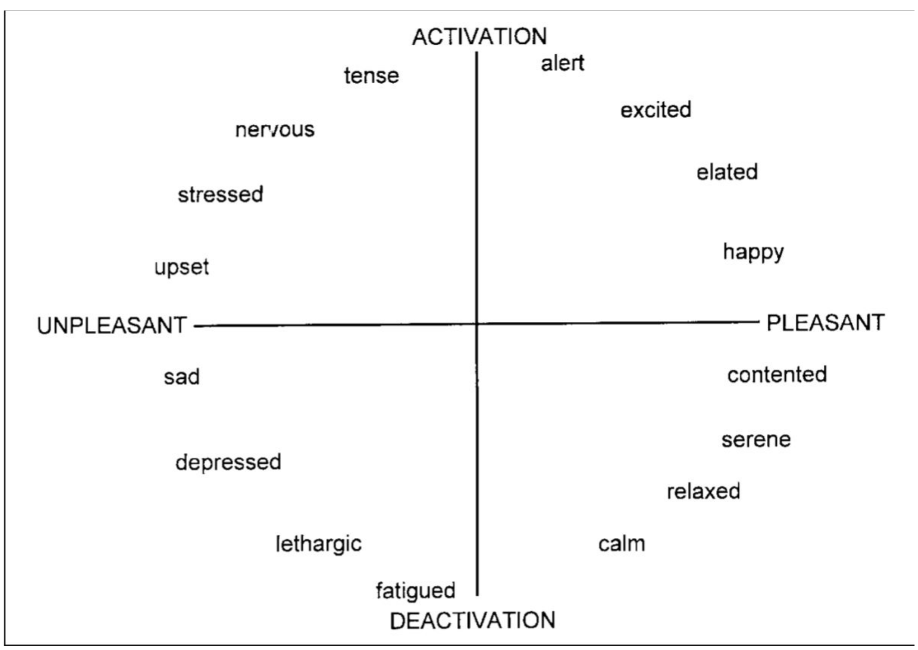


Fig.1: *Schema for the two-dimensional structure of emotion. Adapted from (Feldman Barrett and Russell, 1999).*

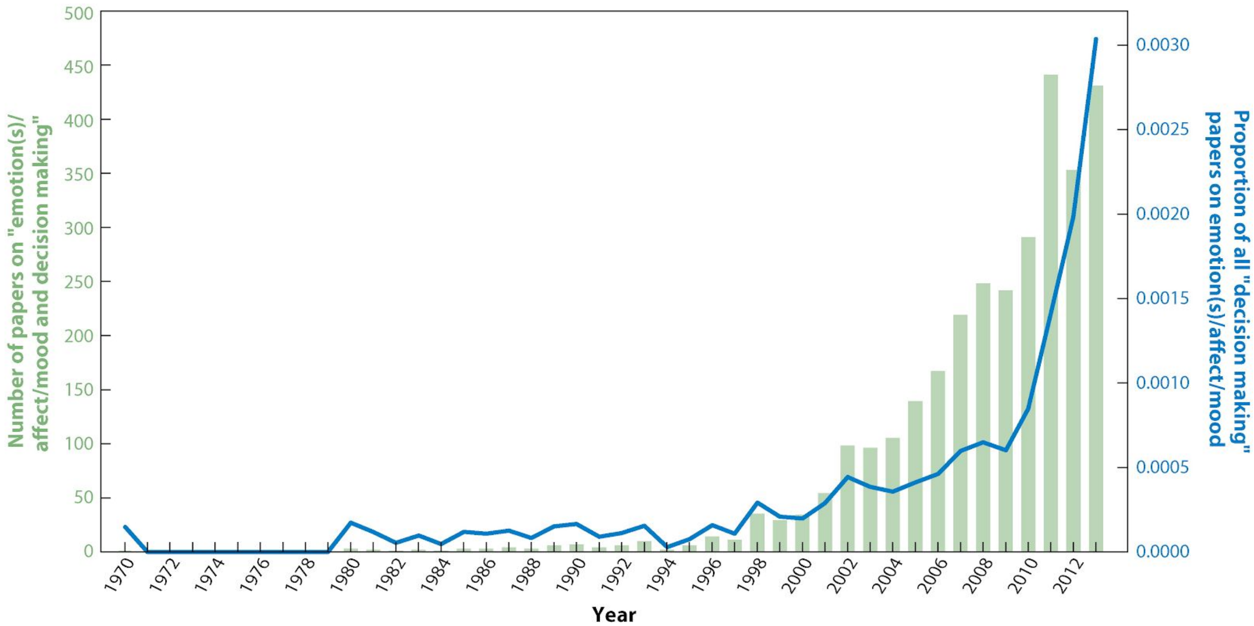
Arousal exerts a well-known effect on the acuity and content of free recall, like “flashbulb memories”, as coined by Brown and Kulik, in which memory for traumatic, exciting, or otherwise emotionally arousing events is heightened (Brown & Kulik, 1977). This can be seen in the vast amount of replication litterature; for instance where people show a heightened ability for recall when it comes to natural disasters and injuries (Bahrick & Levitt, 2001; Bohannon, 1988; “Children’s Memory for Traumatic Injury”,nd). Moreover, evidence from the lab of Micah Allen and others suggests that both pre-stimulus and stimulus-evoked arousal can influence both decision criterion and confidence in decision (Allen et al., 2016; de Gee et al., 2017; Delfini & Campos, 1972).

In regard to valence, a review on the interaction between metamemory and valenced events outline that positive emotions generally show a consequent boost in cognitive performance (Efklides, 2006), but the literature also suggests that people’s confidence in positive events is often higher than actual memory performance - a phenomena coined “optimism bias” (Sharot, 2011). And it is not completely straight forward, as other publications hint at how negative emotions such as worry can also improve meta-cognitive performance (Massoni, 2014).

With this in mind, a novel study has also looked at the interaction between physiological states and metacognition, as seen in a publication where metacognition was modulated by the interaction between the phase of heartbeats and presentation interval of stimuli. Here memory for words presented around systole was lower relative to words presented at diastole, giving clue to how emotions might be able to affect cognitive performance through its effects on the parasympathetic nervous system (Garfinkel et al., 2013). Thus, leveraging Oximetry one can tap into bodily and automatic factors that might affect metacognition.

Together, these studies show contradictory results that make it challenging to draw clear conclusions about emotions influence over memory. However, all the studies hint at some underlying systematicity to how emotions affect both accuracy of recall and confidence in memory. Thus, in this study the aim is to get a clearer understanding of how salient valenced and emotionally charged stimuli affect metamemory.emotionally charged or valenced information.

Putting this work into perspective, it can be seen that it is part of a larger research trend of investigations into how emotions affect cognition, as exemplified in the rise in 'share of publications' now touching on the topic of emotion, in the field of Decision-making (Fig.2).



Lerner JS, et al. 2015.
Annu. Rev. Psychol. 66:799–823

Fig.2: *Number of scholarly publications on emotion(s)/affect/mood and decision making from 1970-2013. Adapted from (Lerner, Li, Valdesolo, & Kassam, 2015).*

The general research landscape is recognising that humans are strained by their imperfect abilities of recollection and moreover, imperfect abilities to perceive these shortcomings. It is an important topic, as memory slip-ups can hold grave consequences, as when people hold unrealistically optimistic views about their own health risks (Shepperd, Pogge, & Howell, 2017). More direly, one can also imagine its consequences in eye witness testimonies, where memories and self-monitoring of memories are exclusively relied on.

Moreover, there exists substantial variation in metacognitive performance (Allen et al., 2017). And poor metacognitive abilities can hold adverse real-life consequences and in the extreme cases lead to systematic biases towards bad decisions, like continuing to follow bad behaviours while being blind to its adverse outcomes, as seen in individuals with Obsessive Compulsive Disorder. (Hauser et al., 2017)

So, in this study we'll examine emotions as a key predictor of influence on metamemory to try and help clarify the specific interactive role between the two, so that in time this knowledge help us develop better metacognitive aids, which already show great promise (Batha & Carroll, 2007; Najmaei, 2016), and ultimately help us fine-tune our behaviour as the emotional people we truly are.

Methodological

This bachelor thesis was registered at Open Science Foundation, as a registered report, where a detailed description of methods and analysis were thought out before conducting the study. This is a good practice as it avoids bad scientific practices such as p-hacking and cherry picked results, as has been critiqued generally in the social sciences ("Scientific Utopia: II —Brian A. Nosek, Jeffrey R. Spies, Matt Motyl, 2012," n.d.)

This project aims to find experimental evidence for interactions between emotional states, accuracy of recall and confidence in recalls (metamemory) by adapting a pre-existing memory task (Li Yan McCurdy et al., 2013).

Metacognition, the measure of one's ability to reflect on own performance, will be tested on correct or incorrect responses. Common methods for analysing metacognition uses statistical correlation coefficients (e.g., Pearson's r) on confidence and accuracy directly to calculate the degree of association directly. But this leaves no option to separate the correct-to-error distribution of these associations, as it excludes measures of both sensitivity and metacognitive sensitivity.

We'll therefore leverage signal detection theoretic (SDT) measures of 'sensitivity' (d') (how one fair in differentiating between correct and incorrect trials), metacognitive sensitivity (md') (how well one's confidence ratings discriminate correct from incorrect trials) and efficiency (mratio) (one's metacognitive sensitivity at different levels of task performance - so, whether variations of confidence match the variation of performances) for stimuli on the emotional dimensions of Arousal {Low,High} and Valence {Negative,Positive} (Maniscalco & Lau, 2014; *Metacoglab/HMeta-d*). This will enable unbiased measures of metacognition that control for perceptual differences (Fleming & Lau, 2014).

For our principal hypotheses tests, we will estimate metacognition variables using the Maximum Likelihood Estimation technique developed by (Maniscalco and Lau, 2012). This approach is currently the most widely used in the field, and has been shown to provide reasonably good model fits when applied to tasks with more than 100 trials per condition. However, a Hierarchical Bayesian model of metacognition (HMeta-d) has recently been shown to have improved estimation even in designs with as few as 20 trials per block. Therefore, to maximize statistical power and compare these methods, we will also test for between-task correlations on metacognitive efficiencies (m-ratio) estimated within the HMeta-d (Fleming, 2017; Mazancieux et al., 2018). The model parameters were sampled using Markov-Chain Monte-Carlo algorithms in JAGS (http://mcmc-jags.sourceforge.net). MCMC estimates are generally reliable for parameter specification robust to local minima. Moreover, a wide standard bayesian prior was used for the HMeta-d as it has been found to be sensitive to group differences (Fleming, 2017).

Additionally, this study will also test whether stimulus-evoked heart-rate changes mediate the influence of arousal and valence on metacognition, as valence and arousal can be predicted from cardiac recording even during short periods of time (Valenza, Citi, Lanatá, Scilingo, & Barbieri, 2014). For this purpose, photoplethysmography (PPG) is used to detect variations in transmitted or reflected light to measure the flow of blood in the finger of participants. This method has provided strong consistency in the estimation of instantaneous heart rate as compared to electrocardiography (Madan, Harrison, & Mathewson, 2018) and has been widely used under similar experimental setups investigating parasympathetic nervous system activation.

Hypotheses

Hypothesis testing:

Our principal tests of interest will be 2x2 (Arousal by Valence) repeated measures ANOVAs on average accuracy, d', reaction time (RT), average confidence, md' and mratio will be estimated using a Bayesian model of metacognition, estimated in a fixed-effects fashion (i.e., without hierarchical group priors) (Fleming, 2017). This will enable us to fit standard frequentists models (e.g., 2x2 ANOVA) to the resulting parameters of interest.

Null Hypothesis: No Effect of Emotion on MetaMemory

Under the null hypothesis, no observable effects are expected of stimulus arousal or valence on confidence or metamemory. This would imply that metacognition is insensitive to emotional inputs, even in the presence of altered accuracy. Statistical support for this hypothesis will be calculated using null Bayes factor analyses in JASP, under a default Cauchy prior = 0.70. By convention (Lee & Wagenmakers, 2014), only a null Bayes factor < 0.33 will be interpreted as evidence for the null hypothesis, whereas BF > 3 would imply evidence for the alternative hypothesis. Thus, BFs between 0.33 and 3 are considered inconclusive.

Alternative Hypothesis 1A: Arousal-Mediated Improvements in Memory Signal to Noise Ratio are Independent of Valence

Under this hypothesis, a main effect of stimulus arousal is expected for recall accuracy and confidence, with no interaction of stimulus valence. This would suggest that arousal serves to sharpen or clarify the stimulus representation in memory, and that metacognition appropriately accounts for this sharpening. Here heart-rate is expected to mediate the influence of accuracy on confidence, but with no difference between valence conditions. This hypothesis will be supported if there is a significant main effect of arousal on these outcome variables, but no main effect or interaction of valence.

Alternative Hypothesis 1B: Valence Bias in MetaMemory

Under this hypothesis, a main effect of stimulus arousal is expected for recall accuracy and confidence. However, the expectation for this effect is dependent on the stimulus valence, such that metamemory for high arousal items will be improved for positive-valence items and reduced for negative-valence items, confirmed via a post hoc pairwise t-test. Under this account, the sharpening effects of arousal on stimulus representation interact with the biasing influence of stimulus valence. Here it is additionally expected that heart-rate mediates the influence of accuracy on confidence in a valence-dependent interaction.

Materials and methods

Participants

Thirty-five participants, who were recruited through local advertisements, took part in the experiment at Skejby Hospital, Aarhus. As it is an exclusively within-subject design, our target sample was N=30 participants, which should deliver sufficient power to detect medium or larger effect sizes. However, to ensure a robust estimate of our behavioural and physiological effect while accounting for potential missing data (due to e.g. poor performance or technical issues with recording) A larger sample of 35 participants was recruited (26 Females) aged 18–26 (M = 21, SD = 1.9) in total.

All participants had normal or corrected to normal vision and were fluent in English. Participants received a monetary compensation of a 100 DKK per hour and the estimated total duration of the test session was 1,5 hour (150 DKK). Participants were also proposed to complete a post-test valence and arousal ratings survey for an additional 50 DKK. 100% (35/35) of the participants accepted to complete this part.
Informed consent was obtained from all participants, and all procedures were conducted in accordance with the Declaration of Helsinki and with approval from the Danish Neuroscience Centre's (DNC) Research Ethics Committee.

Material

The stimuli were 1200 English words selected from the "Affective Norms for English Words" (ANEW) database (Bradley, Lang, Bradley, & Lang, 1999), with valence and arousal ratings as initially measured among a population of English students. Importantly, this database has also been normed in Spanish (Redondo, Fraga, Padrón, & Comesaña, 2007) and Italian (Montefinese, Ambrosini, Fairfield, & Mammarella, 2014) populations, showing good overall consistency in European samples.

The upper and lower tertiles of the valence and arousal word-distribution were used, where words belonging to the neutral intermediate parts were excluded, resulting in 4 distincts subsets of extreme values. Half of the words selected were normed as 'Positive Valence' and the other half as 'Negative Valence'. Among these two subsets, half of the words were also normed as 'High Arousal' in the ANEW database, and the other half was rated as 'Low arousal', leading to four groups of words: 1. Positive-Valence/Low Arousal, 2.Positive-Valence/High-Arousal, 3. Negative-Valence/Low Arousal, 4. Negative-Valence/High-Arousal, each containing 300 words.

Finally, each of these subsets (300 words) were again divided into three equal parts, corresponding to the three 'Learning Time' conditions (LT) (Fig.4A), being either 30, 60 or 90 seconds. Among the 100 resulting words, 50 were then used during the learning phase of the metamemory task, and the 50 remaining were used as distractor words during the test phase. The learning and distracting categories were randomly selected (Fig.4A).

Experimental setup

Overview

The meta-memory experiment had participants memorize 50 words followed by a cued-recall task. The cued-recall task included a two-alternative forced-choice judgment task (2AFC) for item pairs, to discriminate which had been seen previously in the memorized list. After each discrimination, participants rated how confident they were that this 2-AFC judgment was correct using a 7-point Likert scale. This task was adapted from previous studies (L. Y. McCurdy et al., 2013) and was divided into 12 blocks, averaging on a 8 minutes length depending on the participant's performances. Each block consisted of the learning and test of a word list belonging to one of the experimental conditions (Valence {*Negative*, *Positive*}, Arousal {*Low*, *High*}, LT {*30*, *60*, *90*}). Short breaks were interleaved in between each block and the participants could resume the experiment by pressing the space bar in the keyboard. Physiological signals (pulse oximetry) were monitored throughout the experiment with an Oximeter.
In order to validade the ANEW valence and arousing ratings for our sample (1200 words), in the second component of the experiment (Post-Experiment) participants provided their rating for valence and arousal for the words used in the original ANEW survey in a web-based version of the original procedure used in the original ANEW survey (Bradley et al., 1999).

Metamemory Task

Participant instruction and training

Participants were instructed on the meta-memory task. During the briefing the particpants were equipped with an Oximeter on their left middle finger and told that they should keep their hand lying still on the table next to the computer keyboard during the experiment, in order to limit recording artefacts.
All participants completed a brief training protocol of the meta-memory task, introducing them to the experimental design. This protocol included a trial block; remembering 50 words and completing 50 discrimination trials with confidence ratings.

Learning phase

All the 12 experimental blocks started with a learning phase, where the 50 target words were presented to the participant. The words appeared on the screen in the form of a table containing five columns with ten rows each (Fig.3A). The ordering of the words inside these tables remained constant across participants. The list remained on the screen for a period of {*30*, *60*, *90*} seconds depending on the LT condition, after which it automatically switched to the testing phase.
The task was adapted so that each block alternated pseudo-randomly between each level of our 2 by 2 factorial design, with the factors: Valence {*Negative*, *Positive*}, Arousal {*Low*, *High*}, ensuring that highly arousing conditions, for instance, were systematically interleaved by lowly arousing conditions. The block order was also pseudorandomized and counterbalanced across participants (Fig.3A).

Testing phase

During the testing phase, the 50 target words were successively presented to the participant together with a distracting word to the left and right of a fixation cross (Fig.3A). Here, the distractor and target words were selected from lists of 50 words with same arousal and valence levels. Each distractor word was presented only once during this procedure and the distractor lists were counterbalanced across participants. In this way, all words served as either distractor or target, depending on the participant.
The responses were provided using the *right* or *left* keys of the keyboard. After each discrimination, participants rated how confident they were that this choice was correct using a 7-point Likert scale using the *right, left* (Selection) and *down* (Submission) keys of the keyboard. Both for the decision and the confidence reports, the responses had to be provided after 1.5 seconds to avoid errors, and within 3 seconds (Fig. 3A).

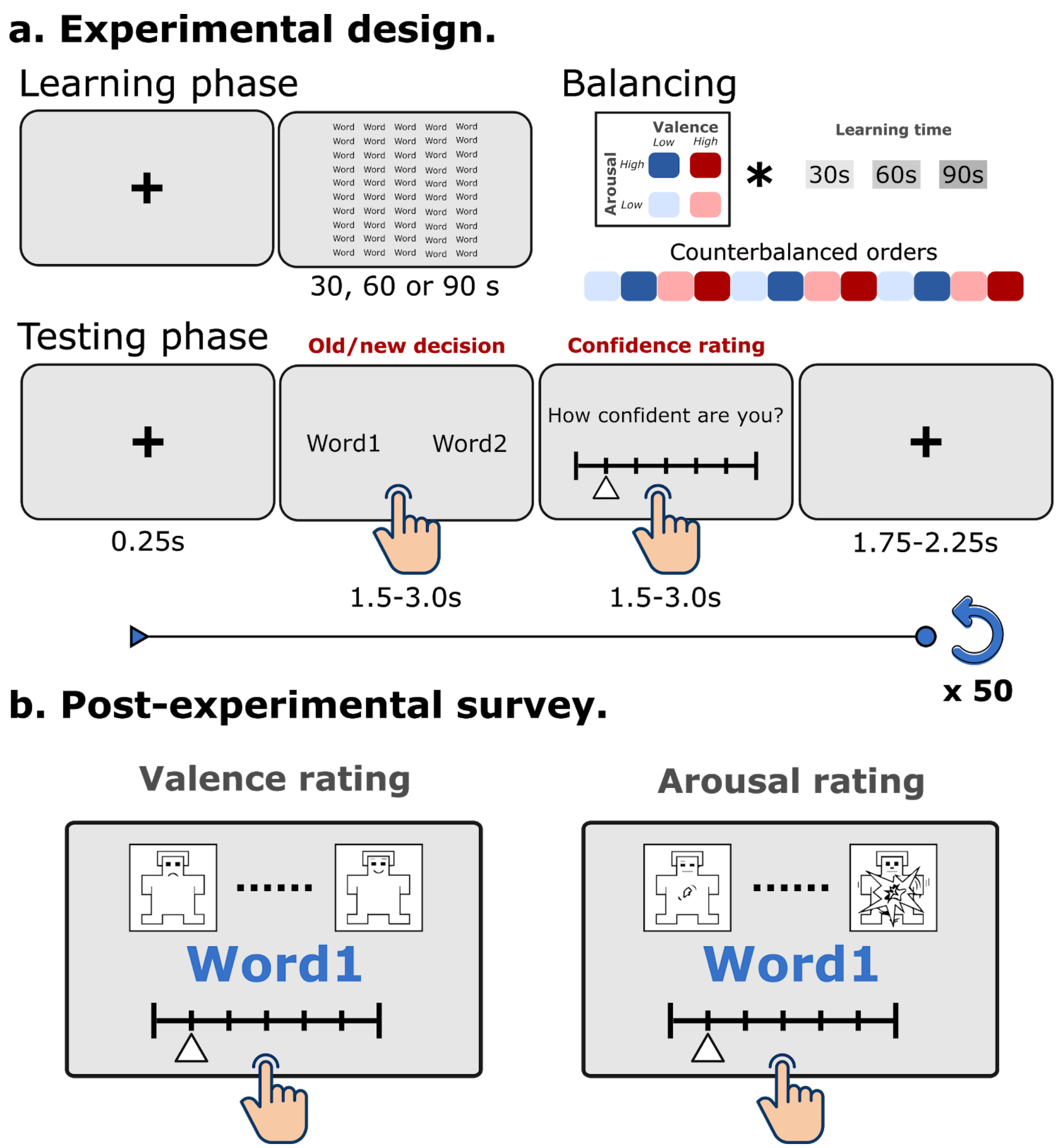


Fig.3: A. *Experimental design. The metamemory task contained 12 experimental blocks, each block consisted of a learning phase and a testing phase for the 50 words. After this procedure, the participants provided in a distinct task their valence and arousal rating for the 1200 words used in the main task (600 target and 600 distractors).* ***B.*** *Post-experimental survey. Participants completed a short subject visual analog scale rating of valence and arousal for the words seen in the experimental design. This was done in a web-based version of the original procedure used in the original ANEW survey. It included a 9-point Likert scale, the word being rated and the original drawings of SAM, showcasing what is meant with Arousal and Valence, all of which were used in the original ANEW survey.*

Physiological monitoring

During the entire experimental procedure, the heart rate of the participants was monitored using the Nonin 3012LP Xpod USB pulse oximeter together with the Nonin 8000SM 'soft-clip' fingertip sensors (https://www.nonin.com/). The sensor was attached to the index finger of the participant's left hand. Pulse oximetry use photoplethysmography (PPG), the detection of variations in transmitted or reflected light to measure the flow of blood. This method has provided strong consistency in the estimation of instantaneous heart rate as compared to electrocardiography (Madan, Harrison, & Mathewson, 2018) and has been widely used under similar experimental setups.

Post-Experiment

For the second component of the experimental design, the participants were asked to complete an at-home post-experiment survey, to collect subjective ratings for the words used during this experiment. Participants completed a short subject visual analog scale rating of the valence and arousal in a web-based version of the original procedure used in the original ANEW survey (Bradley et al., 1999). This part was about 30 to 60 minutes long and was carried out using Pavlovia (https://pavlovia.org/.org), an online platform for running PsychoPy experiments (Peirce et al., 2019). Each word was presented twice with a 9-point Likert scale for valence and arousal in turn. These scales were complemented with pictures of the original drawings of the Self-Assessment Manikin (Bradley & Lang, 1994), which were also used in the original ANEW survey (Bradley et al., 1999). Here, there was no time limit for responses. The consistency between the ANEW ratings and the ratings provided by the participants is depicted in (Fig.4B).

This was carried out in order to confirm that the selection based on the ANEW rating was consistent with the evaluation of the participants, because a full-blown double translation and revalidation of the ANEW library in Danish was not possible for this study.
Nevertheless, ANEW has been normed in Dutch, Spanish and Italian populations, showing good overall consistency in European samples (Montefinese, Ambrosini, Fairfield, & Mammarella, 2014; Moors et al., 2013; Bradley, Lang, Bradley, & Lang, 1999; Redondo, Fraga, Padrón, & Comesaña, 2007). All participants did provide new ratings for experimental stimuli.

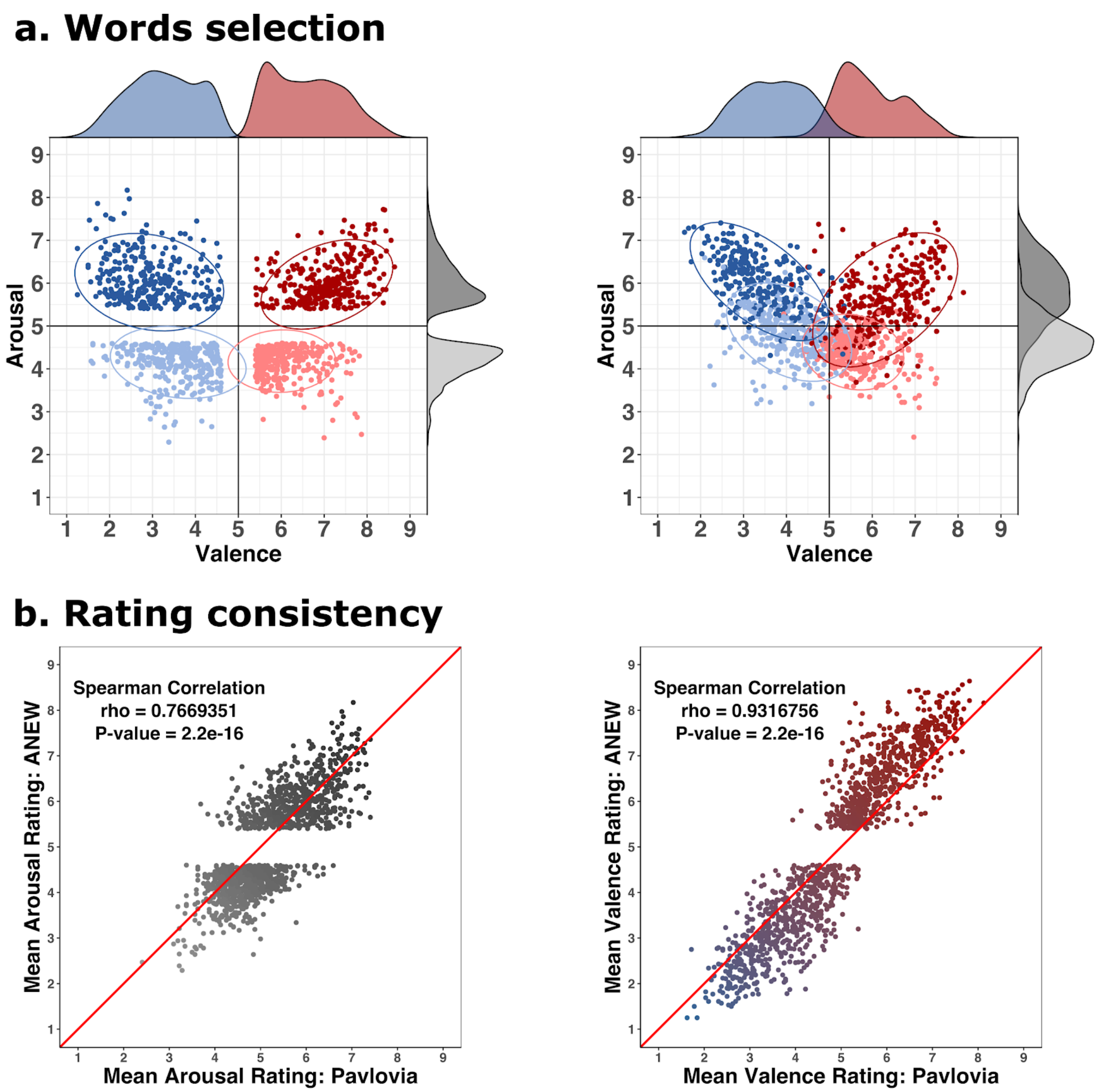


Fig.4: A. *Word selection. We selected the words used in the metamemory task in the highest and lowest tertile of the Valence and Arousal ratings as estimated by the ANEW database, thereby excluding the points belonging to one of the intermediate categories.* **B.** *Concordance of the subjective rating with selection. The independent rating provided by the ANEW database was compared to the actual ratings provided by the participants after the main procedure.*

Analysis

OSF

Prior to data collection this project was put up on the Open Science Foundation (OSF) where all exclusion criteria stated below were written in advance. (Link to OSF: https://osf.io/9awtb)

Behavioural

For each participant the OSF pre-registered cleaning procedure was followed (exclusion criteria and planned transformations). All trials in the meta-memory experiment with a RT faster than 100ms were excluded. Moreover, any trials with RTs greater than 3 SD away from the median RT were removed. Participants with missing data (i.e. data loss due to technical failures) or insufficient data (fewer than 50 trials in total per condition) were also excluded.
Inherent in the Python script that ran the experiment was a bug that made late anwsers/missed trials (Within the 3 sec window) get accepted, thereby overwriting the next discrimination task choice. Therefore all NA trials were excluded as well as the following trial.

Following these trial level exclusions, each subjects performance on the 2AFC task was calculated using SDT measures of sensitivity (d'), metacognitive sensitivity (md') metacognitive efficiency (mratio), as well as mean accuracy, mean confidence and mean RT separately for each of the four conditions. These measures collectively summarized the memory performance and metacognition.

Finally, outliers in task performance were detected by inspecting boxplots for reaction time, d', and confidence across all subjects; subjects whose values for any of these variables on any condition are greater than 1.5 times the interquartile range were be excluded from the analysis.

From the original 35 participants, a total of 29 participants passed the exclusion criteria and were further analyzed. This sample size is just below the target sample (30).

The preprocessing of the behavioural data was carried out using custom R scripts (https://github.com/sebsebar/MetaCognition) using R Studio (1.2.5019) and the R software (R 3.6.1) on Mac OS.

All SDT measures (d', md', mratio) were derived from the HMeta-d model (*Metacoglab/HMeta-d*, 2014/2019), adapted in R using custom scripts (https://github.com/sebsebar/MetaCognition), run in the non-hierarchical mode to allow frequentist analysis of the resultant parameters.

Physiological monitoring

The preprocessing and statistical analysis of the pulse oximetry recordings also followed the OSF pre-registered cleaning procedure. Firstly the PPG signal was smoothed to a 1000 Hz using a cubic spline interpolation. A moving average was then used to remove high-frequency noise and control clipping artefacts (window of 0.2s). Peaks were then detected using a moving average with a standard deviation (SD) threshold. All continuous segments higher than the mean plus 1 SD (window of 0.75s), were labelled the 'local maximum' and thereby also peak. This peak corresponds to the higher level of blood oxygen measured as in the middle finger of the participant and is connected to the T wave found in the ECG recording. Using this peak-to-peak interval produces similar Heart Rate Variability (HRV) metrics as compared to the R-R interval used in traditional HRV research (SAYKRS, 1973).

The peak-to-peak interval was then used to extract interbeat intervals (IBI) time series and derive HRV metrics. Controlling for outliers was performed by removing extremely high (>2500 ms) and extremely low (<300ms) intervals. Intervals above and below this can't come from a normally operating heart and were thus removed either by suppression of the peak or by adding one or more peaks separated by equal intervals. Number of peaks needed was assessed by minimizing the total SD of the IBI time series.

Additionally, outliers were detected and corrected using the generalized extreme studentized deviate test from the Anomaly Detection Toolkit (https://arundo-adtk.readthedocs-hosted.com/en/stable/).

Finally, the 'Root Mean Squared Standard Deviation' (RMSSD) was extracted, which is a common HRV metric according to standard guidelines and reflects the vagally mediated activity of the parasympathetic nervous system ("Heart rate variability," 1996; Shaffer & Ginsberg, 2017).

For consistency, results were only reported for the 29 participants that were included in the previous behavioral section.

Post-Experiment

In the second component of the experimental design, the participants did an at-home post-experiment survey as to compare participant ratings with the ANEW ratings. Participants were excluded on the basis of boxplot inspections; excluding participants who didn't use the 9-point likert scale (Fig.3B) on the grounds that having more than half of their ratings being 5 (Neutral) for either the valence or arousal ratings suggested a disregard for the instructions to use the scale.

After the exclusions, 31 participants from the initial 35 were included in further analysis.

The non-parametric Spearman rank correlation test was used to compare the post-experiment word ratings to the ANEW, as it does not carry any assumptions about the distribution of the data and is therefore best suited for ordinal data.

Statistical

Primary Analysis

The primary analysis consists of a 2x2 Repeated Measures ANOVAs on d', md', m-ratio, average confidence, average accuracy and mean RT. These were conducted in JASP using frequentist statistical tests, alpha level = 0.95. In case of an interaction a multiple comparison Bonferroni corrected post hoc tests was performed to give a conservative estimate of the interactive effects.
If an ANOVA solely rendered null-effects an equivalent Bayesian test was run under the default cauchy prior = 0.707 in order to assess the evidence for the null hypothesis, as proposed by Jeffreys on this part of the analysis (1961, see also Rouder et al. 2009). Here only BFs > 3 or < 0.33 were considered informative as outlined in the pre-registration.

Exploratory Analysis (Learning Time)

Additionally, to explore the influence of LT on the SDT estimates, the behavioural analysis/pre-processing and task-level exclusions for each of the 3 LT {*30*, *60*, *90*} (Fig.3A) was carried out again, now leaving out the excluding outliers in task performance, as only 14 participants would make this cut, outlined in the OSF exclusion criteria, after dividing the data up into 3 parts.
Thus, the primary analysis was run again, but now dependent on LT.

The bayesian and frequentists models were implemented using the JASP software (https://jasp-stats.org/) version 0.11.1. All SDT measures (d', md' and mratio) were derived from the HMeta-d (Fleming, 2017) implemented in R (https://github.com/metacoglab/HMeta-d), run in the non-hierarchical mode to allow frequentist analysis of the resultant parameters. The R scripts implementing the modeling part can be retrieved here: (https://github.com/sebsebar/MetaCognition). Everything was run on Mac OS.

Results

Primary Analysis

Discrimination

Accuracy. A Valence x Arousal ANOVA showed no effect of Arousal ($F_{(1,28)} = 0.017$, $\eta_p 2 = 0.001$, $p = 0.897$), but an effect of Valence ($F_{(1,28)} = 10.093$, $\eta_p 2 = 0.265$, $p = 0.004$) with an interaction between these two factors ($F_{(1,28)} = 6.698$, $\eta_p 2 = 0.193$, $p = 0.015$).
Planned comparisons on interactions revealed that participants had a higher Accuracy, on average, of stimuli with 'Positive Valence' than with 'Negative Valence' under the condition of 'Low Arousal' ($t_{(28)} = -4.092$, $p < 0.001$, $d = -0.760$, two-tailed).

d'. A Valence x Arousal ANOVA showed no effect of Arousal ($F_{(1,28)} = 0.019$, $\eta_p 2 = 0.001$, $p = 0.891$), but an effect of Valence ($F_{(1,28)} = 10.672$, $\eta_p 2 = 0.276$, $p = 0.003$) with no interaction between these two factors ($F_{(1,28)} = 4.30$, $\eta_p 2 = 0 .133$, $p = 0.047$).
Planned comparisons on interactions revealed that participants had a higher sensitivity (fewer misses and false alarms) on average, of items with 'Positive Valence' than with 'Negative Valence' under the condition of Low Arousal ($t_{(28)} = -3.787$, $p = 0.002$, $d = -0.703$, two-tailed).

Accuracy RT. A Valence x Arousal ANOVA showed no effect of Arousal ($F_{(1,28)} = 0.4086$, $\eta_p 2 = 0.127$, $p = 0.053$), but an effect of Valence ($F_{(1,28)} = 4.657$, $\eta_p 2 = 0.143$, $p = 0.040$), with an interaction between these two factors ($F_{(1,28)} = 6.890$, $\eta_p 2 = 0 .197$, $p = 0.014$).
Planned comparisons on interactions revealed that participants had a faster RT on the 2AFC task, on average of items with 'Positive Valence' under the 'Low Arousal' condition compared to the 'High Arousal' condition ($t_{(28)} = 3.312$, $p = 0.010$, $d = 0.615$, two-tailed) and a faster RT on the 2AFC task, on average of items with 'Positive Valence' compared to 'Negative Valence' under the of 'Low Arousal' condition ($t_{(28)} = 3.335$, $p = 0.001$, $d = 0.619$, two-tailed).

MetaCognition

Confidence. A Valence x Arousal ANOVA showed no effect of Arousal ($F_{(1,28)} = 0.177$, $\eta_p 2 = 0.006$, $p = 0.677$), but an effect of Valence ($F_{(1,28)} = 13.760$, $\eta_p 2 = 0.329$, $p < 0.001$), with no interaction between these two factors ($F_{(1,28)} = 3.772$, $\eta_p 2 = 0 .119$, $p = 0.062$). Meaning that 'Positive Valence' stimuli generally provoked higher confidence ratings.

Meta d'. A Valence x Arousal ANOVA showed an effect of Arousal ($F_{(1,28)} = 4.591$, $\eta_p 2 = 0.141$, $p = 0.041$) and of Valence ($F_{(1,28)} = 15.473$, $\eta_p 2 = 0.356$, $p < 0.001$) , though no interaction between these two factors ($F_{(1,28)} = 1.451$, $\eta_p 2 = 0.049$, $p = 0.238$). Meaning that 'Positive Valence' and 'Low Arousal' stimuli generally provoked higher metacognitive sensitivity in their confidence ratings, meaning that the participants felt more confident in their judgments.

Mratio. A Valence x Arousal ANOVA showed no effect of Arousal ($F_{(1,28)}$ = 0.607, η_p^2= 0.021, p = 0.443), and no effect Valence ($F_{(1,28)}$ = 1.441, η_p^2 = 0.049, p = 0.240), as well as no interaction between these two factors ($F_{(1,28)}$= 0.407, η_p^2 = 0 .014, p = 0.529).
As there were no effects for any of the predictors a Bayesian t-test framework was used, according to the OSF pre-registration. The predictor with the highest bayes factor is 'valence': (BF10 = 0.348) indicating that the evidence for the Null hypothesis H0 is approximately 2.87 times more likely to occur than H1, given the data. But, keeping in mind our criteria of only considering BFs > 3 or < 0.33 as informative, this result is inconclusive.

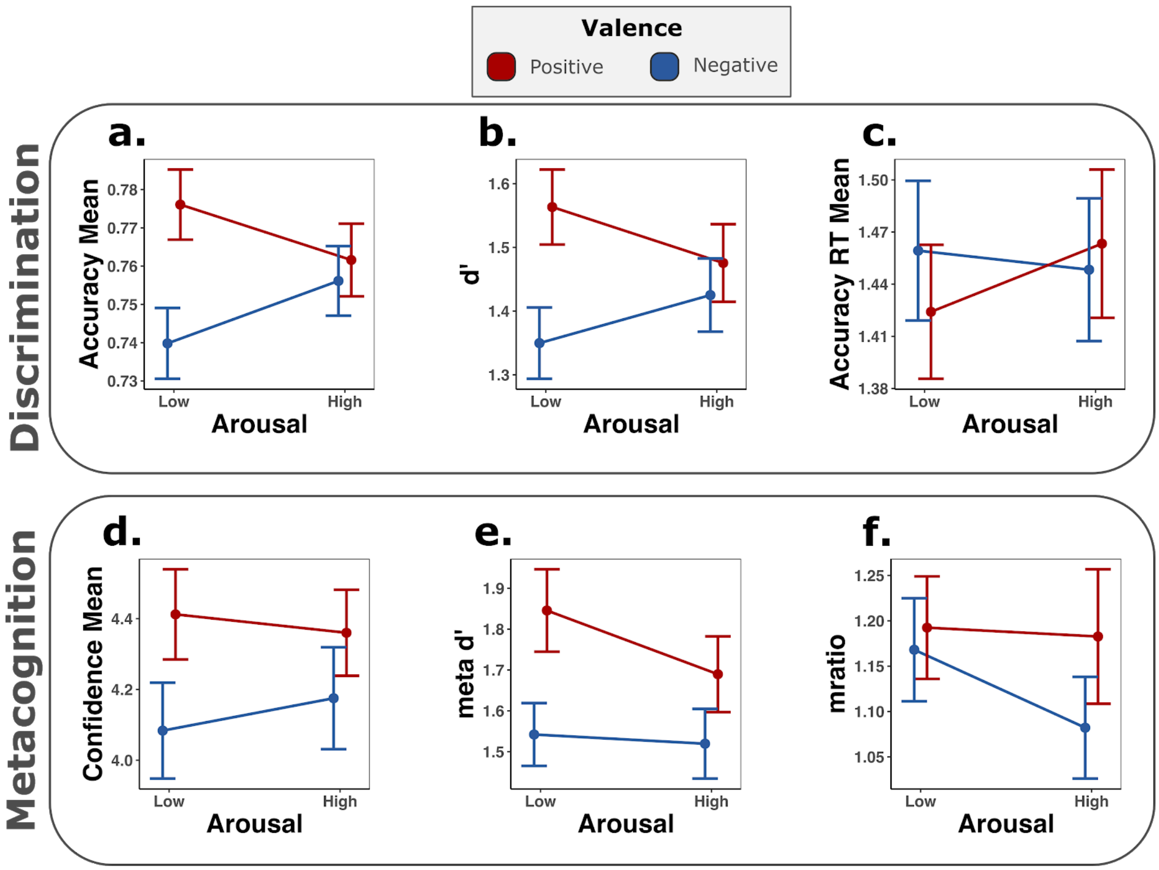


Fig.5: *Results of primary analysis.* **A.** *Mean Accuracy over different levels of Arousal and Valence,* **B.** *d' over different levels of Arousal and Valence,* **C.** *Accuracy RT over different levels of Arousal and Valence,* **D.** *Mean Confidence over different levels of Arousal and Valence,* **E.** *Meta d' over different levels of Arousal and Valence,* **F.** *Mratio over different levels of Arousal and Valence.*
The error bars represent standard error of the mean (SEM).

Physiological monitoring

Heart Rate

A Valence x Arousal ANOVA showed an effect of Arousal ($F_{(1,25)}$ = 6.33, η_p^2= 0.20, p = 0.018), and an effect Valence ($F_{(1,25)}$ = 8.72, η_p^2 = 0.25, p = 0.006), but no interaction between these two factors ($F_{(1,25)}$= 0.71, η_p^2 = 0 .028, p = 0.40). Meaning that 'Positive Valence' and 'High Arousal' stimuli generally provoked higher heart rates as shown in BPM.

Time Domain Analysis

A Valence x Arousal ANOVA showed no effect of Arousal ($F_{(1,25)}$ = 0.70, η_p^2= 0.027, p = 0.40), but an effect of Valence ($F_{(1,25)}$ = 19.38, η_p^2 = 0.43, p < 0.001), but no interaction between these two factors ($F_{(1,25)}$= 0.86, η_p^2 = 0 .034, p = 0.36). Meaning that 'Positive Valence' stimuli generally provoked lower RMSSD values.

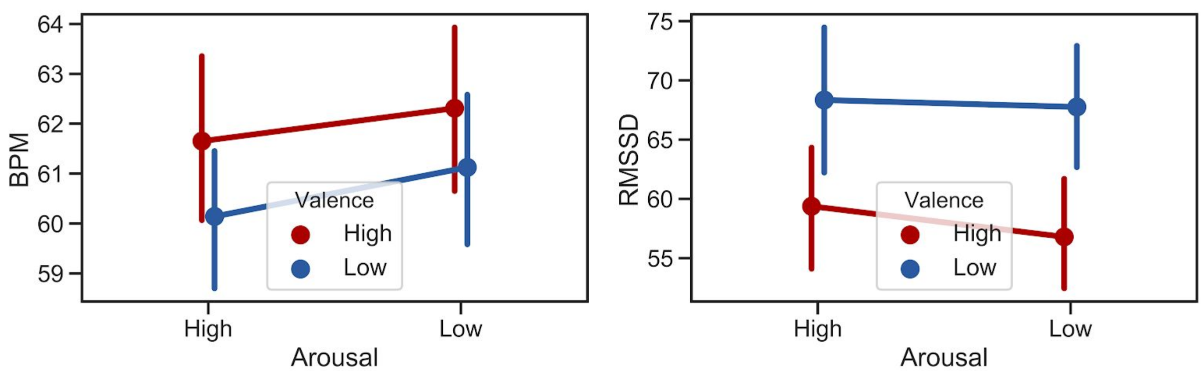


Fig. 6: *Physiological monitoring.* **A:** *Average heart rate (BPM) over different levels of Arousal and Valence* **B:** *RMSSD over different levels of Arousal and Valence. The error bars represent standard error of the mean (SEM).*

Post-Experiment

A non-parametric Spearman rank correlation was performed on the ratings from the ANEW database and the ratings assembled through the post experiment survey, showing strong effects (Akoglu, 2018) for both Arousal (rs(30) = .77, p < .001) and Valence (rs(30) = .93, p < .001) (Fig.4B).

Exploratory analysis (Learning Time)

A full-blown exploratory analysis is beyond the scope of this thesis, compared to the eventual publication. Accuracies and Confidence ratings are the only effects to be evaluated.

Discrimination

Accuracy. A Valence x Learning Time ANOVA showed a strong effect of Learning Time ($F_{(1,28)}$ = 95.657, $\eta_p 2$= 0.319, p < 0.001) a Valence ($F_{(1,28)}$ = 8.067, $\eta_p 2$ = 0.012, p =0.008) with no interaction between these two factors ($F_{(1,28)}$= 4.251, $\eta_p 2$ = 0.12, p = 0.200). Meaning that 'Positive Valence' stimuli generally provoked higher Accuracy than 'Negative Valence' and that higher LT also facilitated high Accuracy.

MetaCognition

Confidence. A Valence x Learning Time ANOVA showed a strong effect of Learning Time ($F_{(1,28)}$ = 77.421, $\eta_p 2$= 0.265, p < 0.001), and an effect of Valence ($F_{(1,28)}$ = 7.240, $\eta_p 2$ = 0.016, p =0.012), but no interaction between the two factors ($F_{(1,28)}$= 1.600, $\eta_p 2$ = 0.006, p = 0.211). Meaning that 'Positive Valence' stimuli generally provoked higher confidence ratings than 'Negative Valence' and that higher LT also facilitated higher Confidence Ratings.

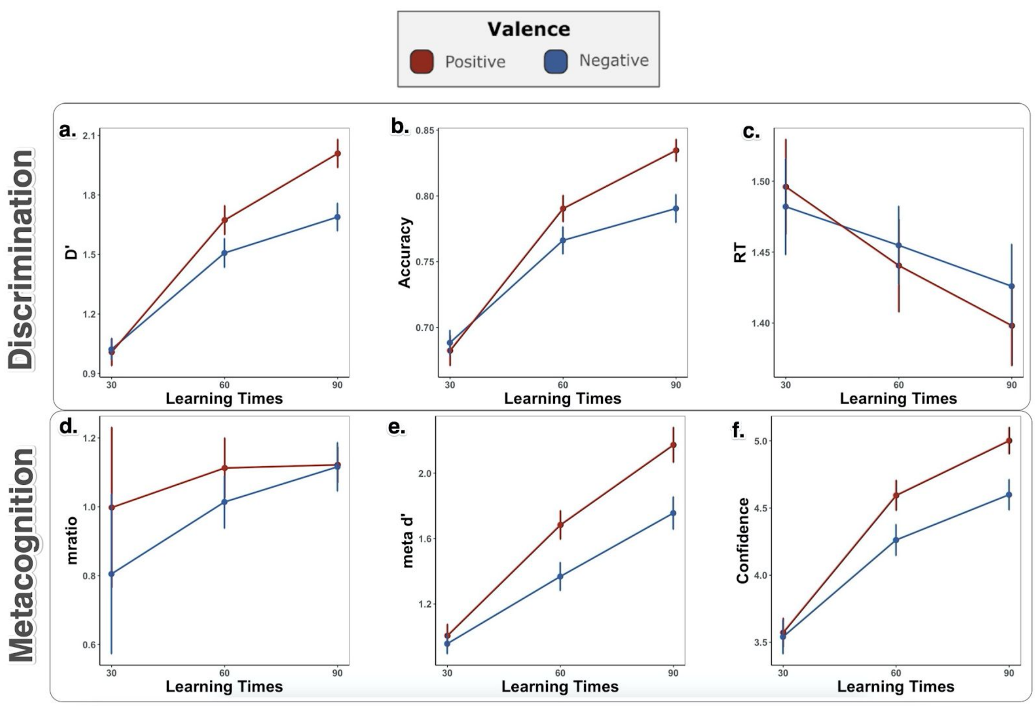


Fig.7: Results of exploratory analysis. **A.** *d' over LearningTime over different levels of Valence,* **B.** *Mean Accuracy over LearningTime over different levels of Valence,* **C.** *Accuracy RT over LearningTime over different levels of Valence,* **D.** *mratio over LearningTime over different levels of Valence,* **E.** *Meta d' over LearningTime over different levels of Valence,* **F.** *Mean Confidence over LearningTime over different levels of Valence.*
The error bars represent standard error of the mean (SEM).

Discussion

The results indicate that salient valenced stimuli regulates memory, as it showed an effect on all discrimination measures as well as confidence and metacognitive sensitivity, supporting a role for emotions in guiding both general memory performance and metamemory. These findings are nevertheless mostly limited to low the arousal stimuli group for discrimination measures, as clearly seen in the interaction effects of both mean accuracy and d' (Fig.5) implying a diminishing effect of 'Positive Valence' on memory performance when stimuli is 'High Arousal'.

Yet, for the metacognition measures of mean confidence and meta d', 'Positive Valence' stimuli provoked higher ratings/metacognitive sensitivity across Arousal conditions.

Interestingly this study did not show clear a general effect of arousal on discrimination, and only a negative trend for metacognitive sensitivity where 'High Arousal' showed lower meta d' values compared to the 'Low Arousal' condition.

This conflicts with earlier studies that found clear effects of stimulus-evoked arousal (High Arousal) on confidence (et al., 2016; Delfini & Campos, 1972; "Dynamic modulation of decision biases - Europe PMC," n.d.).

Nonetheless, this study showed effects of Arousal as interacting with Valence, implying the importance of Valence, and concurring with earlier reviews (Efklides, 2006) showing that positive emotions generally show a boost in cognitive performance.

In addition to the primary analysis, the physiological monitoring also showed how 'Positive Valence' rendered a positive effect both lowering RMSSD, reflecting a higher vagally mediated activity of the parasympathetic nervous system, and also raising heart rate as seen in BPM (Fig. 6), strengthening the hypothesis that the heart holds a mediatory role in relation to the valence-dependent effects, as seen alike in the work of (Garfinkel et al., 2013).

One drawback of this investigation is that the measure of metacognitive efficiency (mratio), reflecting how variations of confidence match the variation of accuracy, rendered no significant effects. Moreover, it can not be considered a null-findinding as the Bayes factor (BF10=0.348) is still within the limits of what is considered uncertain results (BFs > 3 or < 0.33). This means that no clear conclusions can be drawn about the "optimism bias" for instance, that other studies have found (Sharot, 2011).

The uncertainty here might be a matter of having too noisy estimates of mratio. This could be remedied through running the full hierarchical model adding group-level prior densities to each of the subject-level parameters (Fleming, 2017). This was not done in the primary analysis as to not violate the assumption of independence of the standard frequentists models that were used here.

On a different node, the exploratory analysis found that LT facilitated both higher Accuracy and Confidence (Fig.7). 'Positive Valence' stimuli generally provoked higher Accuracy and Confidence than 'Low Valence', but was not shown to significantly interact with LT. However, eyeballing the plot in

(Fig.7) seem to suggest that LT could indeed be driving a the effect of Valence and thereby the separation between the two Valence groups across levels.

This curiosity indicates that an extension of the study could be warranted. Such interaction is suggestive at best, both because the interaction is not significant, but also because the whole exploratory analysis was not put under the same exclusion-criteria as the primary analysis, and thereby is less reliable.

However, if one redid the study, one could also try having longer LTs, as to test whether the separation of the Valence groups keeps growing. In regard to this, other procedure than the 2AFC might also be leveraged (Fairfield, Mammarella, Palumbo, & Di Domenico, 2015), as for example the 'ease of learning' task that would enable on to add more levels to the Valence stimuli e.g. {Highly Negative, Negative, Neutral, Positive, Highly Positive} as seen in the work of Mazzoni ("Remembering the Grocery Shopping List: Mazzoni n.d.). This could also be interesting for another reason, which is that post-trial confidence-based metacognition may not generalize to all metacognitive judgments (Li Yan McCurdy et al., 2013).

Something noticed during data collection was also that many participants reported using different strategies for memorisation, a redo of the study could potentially also investigate if different strategies would be more or less prone to the effects of salient valenced and emotionally charged stimuli. So by redesigning this study or scaling it up, one could both get a clearer understanding of the Valence mediating memory effects, but possibly also identify memory aid strategies.

As goes for the stimuli control, participant word ratings showed a strong correlation with the original ANEW ratings (Arousal = 0.77 & Valence = 0.93). However, there was also a tendency for centrality, as can be clearly seen with the overlap on the y-axis/Arousal dimension (Fig.4A).

A potential explanation is that unlike the original ANEW study, the participants in this study didn't rate any of the words that were deemed 'neutral' in the ANEW database, as those were not part of the Metamemory task, and therefore not on the online survey. This could have induced a consequently more neutral/conservatory rating, since the baseline of words up for rating was different compared to the original study.

By closer inspection, several other limitations become also apparent considering the ANEW database. Firstly, looking at the rating consistency (Fig.4B), it can be seen that words were more positively rated in the Post-Experiment than the ANEW database for the 'Negative Valence' words and more negatively rated for the 'Positive Valence' words, as seen in the skew from the identity line (Fig.4B). This could indicate that unlike the Spanish (Redondo et al., 2007) and Italian (Montefinese et al., 2014) populations, that showed good overall consistency, Danes differ.

On this note, another problem with the ANEW database was detected. Some word ratings seem peculiar - e.g. having '*pregnant*' in the 'LowValenceHighArousal' group next to words like '*grenade*' and '*wreck*'. Such a rating might not hold outside a limited sample of American Psychology class students.

Nevertheless, the control did show strong correlations between the study's ratings and the ANEW database, supporting the use of the ANEW database for the current investigation.

Conclusion

This study sheds light on the impact of valence and arousal on learning new words. It shows how emotional words impact what subjects remember and how well they remember it. Specifically we see how salient valenced stimuli show a strong positive effect on memory performance and metacognition supporting a role for emotions in guiding confidence and memory performance.

There was support for H1B as we saw a heart-rate mediated influence on the sharpening effects of arousal on accuracy and confidence in a valence-dependent effect/interaction. However, this study found little evidence for an independent effect of arousal on discrimination and metacognition, contrary to prior studies (Allen et al., 2016; de Gee et al., 2017;Delfini & Campos, 1972; "Dynamic modulation of decision biases—Europe PMC," n.d.). Instead effects were found for Arousal mostly dependent on Valence. The evidence for H1A, an arousal-mediated improvements in memory independent of valence, was in this regard not strengthened.

Future work will be to scale up the experiment with longer LT in the 'Learning Phase' as well as include more participants to explore the effects of Valence further. Also, adopting the full hierarchical model of (Fleming & Lau, 2014) would be suggested as to render the metacognitive efficiency (mratio) interpretable. Ultimately, a full-blown replication of the ANEW database in Danish would also be wise moving forward.

References

https://doi.org/10.1523/JNEUROSCI.1890-12.2013

17th and 18th Century Theories of Emotions > Descartes on the Emotions (Stanford Encyclopedia of Philosophy). (n.d.). Retrieved December 4, 2019, from https://plato.stanford.edu/entries/emotions-17th18th/LD2Descartes.html

Allen, M., Frank, D., Schwarzkopf, D. S., Fardo, F., Winston, J. S., Hauser, T. U., & Rees, G. (2016). Unexpected arousal modulates the influence of sensory noise on confidence. *ELife*, *5*, e18103. https://doi.org/10.7554/eLife.18103

Bahrick, L. E., & Levitt, M. J. (2001). *MEMORY FOR A NATURAL DISASTER 331 Food ? Cooking ? Drinking ? Ice ? Refrigerator ?*

Barrett, L. F., & Russell, J. A. (1999). The Structure of Current Affect: Controversies and Emerging Consensus. *Current Directions in Psychological Science*, *8*(1), 10–14. https://doi.org/10.1111/1467-8721.00003

Bohannon, J. N. (1988). Flashbulb memories for the space shuttle disaster: A tale of two theories. *Cognition*, *29*(2), 179–196. https://doi.org/10.1016/0010-0277(88)90036-4

Bradley, M. M., Lang, P. J., Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction manual and affective ratings*.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*(1), 73–99. https://doi.org/10.1016/0010-0277(77)90018-X

Brown, R., & McNeill, D. (1966). The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 325–337. https://doi.org/10.1016/S0022-5371(66)80040-3

Campos-Vazquez, R. M., & Cuilty, E. (2014). The role of emotions on risk aversion: A Prospect Theory experiment. *Journal of Behavioral and Experimental Economics (Formerly The Journal of*

Socio-Economics), *50*(C), 1–9.

Children's Memory for Traumatic Injury—Peterson—1996—Child Development—Wiley Online Library. (n.d.). Retrieved December 4, 2019, from https://srcd.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-8624.1996.tb01902.x?casa_token=JphmJfXRzIcAAAAA%3APTgQjOs2gV9Pa009OrjhqIAhitpZg_h9H5O4ITL93I1bnbyTWzDH_1d_hXWSygITz7gp8_DnV4-FOAg

Damasio: Descartes' Error: Emotion, Reason and the... - Google Scholar. (n.d.). Retrieved December 2, 2019, from https://scholar.google.com/scholar_lookup?title=Descartes%27%20Error%3A%20Emotion%2C%20Reason%2C%20and%20the%20Human%20Brain&author=A.R.%20Damasio&publication_year=1994

Delfini, L. F., & Campos, J. J. (1972). Signal Detection and the "Cardiac Arousal Cycle." *Psychophysiology*, *9*(5), 484–491. https://doi.org/10.1111/j.1469-8986.1972.tb01801.x

Dunning, D., Heath, C., & Suls, J. M. (2016). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest*. Retrieved from https://journals.sagepub.com/doi/10.1111/j.1529-1006.2004.00018.x

Dynamic modulation of decision biases by brainstem arousal systems. - Abstract—Europe PMC. (n.d.). Retrieved December 4, 2019, from https://europepmc.org/article/PMC/5409827

Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, *1*(1), 3–14. https://doi.org/10.1016/j.edurev.2005.11.001

Fairfield, B., Mammarella, N., Palumbo, R., & Di Domenico, A. (2015). Emotional Meta-Memories: A Review. *Brain Sciences*, *5*(4), 509–520. https://doi.org/10.3390/brainsci5040509

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Flavell, J. H., & Wellman, H. M. (1975). *Metamemory*. Retrieved from https://eric.ed.gov/?id=ED115405

Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, *2017*(1). https://doi.org/10.1093/nc/nix007

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00443

Garfinkel, S. N., Barrett, A. B., Minati, L., Dolan, R. J., Seth, A. K., & Critchley, H. D. (2013). What the heart forgets: Cardiac timing influences memory for words and is modulated by metacognition and interoceptive sensitivity. *Psychophysiology*, *50*(6), 505–512. https://doi.org/10.1111/psyp.12039

Heart rate variability: Standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996). *Circulation*, *93*(5), 1043–1065.

Huebner, B., Dwyer, S., & Hauser, M. (2009). The role of emotion in moral psychology. *Trends in Cognitive Sciences*, *13*(1), 1–6. https://doi.org/10.1016/j.tics.2008.09.006

Koriat: Metamemory: The feeling of knowing and its vagaries—Google Scholar. (n.d.). Retrieved December 18, 2019, from https://scholar.google.com/scholar_lookup?journal=International+Journal+of+Psychology&title=Metamemory:+The+feeling+of+knowing+and+its+vagaries&author=A+Koriat&volume=31&issue=3%E2%80%934&publication_year=1996&pages=4611-4611&

Ledoux, J. (1998). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*. Simon and Schuster.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and Decision Making. *Annual Review of Psychology*, *66*(1), 799–823. https://doi.org/10.1146/annurev-psych-010213-115043

Liu, L., Feng, T., Chen, J., & Li, H. (2013). The Value of Emotion: How Does Episodic Prospection Modulate Delay Discounting? *PLOS ONE*, *8*(11), e81717. https://doi.org/10.1371/journal.pone.0081717

Madan, C. R., Harrison, T., & Mathewson, K. E. (2018). Noncontact measurement of emotional and physiological changes in heart rate from a webcam. *Psychophysiology*, *55*(4), e13005. https://doi.org/10.1111/psyp.13005

Maniscalco, B., & Lau, H. (2014). Signal Detection Theory Analysis of Type 1 and Type 2 Data: Meta-d′, Response-Specific Meta-d′, and the Unequal Variance SDT Model. In S. M. Fleming & C. D. Frith (Eds.), *The Cognitive Neuroscience of Metacognition* (pp. 25–66). https://doi.org/10.1007/978-3-642-45190-4_3

Massoni, S. (2014). Emotion as a boost to metacognition: How worry enhances the quality of confidence. *Consciousness and Cognition*, *29*, 189–198. https://doi.org/10.1016/j.concog.2014.08.006

McCurdy, Li Yan, Maniscalco, B., Metcalfe, J., Liu, K. Y., Lange, F. P. de, & Lau, H. (2013). Anatomical Coupling between Distinct Metacognitive Systems for Memory and Visual Perception. *Journal of Neuroscience*, *33*(5), 1897–1906. https://doi.org/10.1523/JNEUROSCI.1890-12.2013

Metacoglab/HMeta-d [MATLAB]. (2019). Retrieved from https://github.com/metacoglab/HMeta-d (Original work published 2014)

Metacognition: Knowing about Knowing—Google Books. (n.d.). Retrieved December 20, 2019, from https://books.google.dk/books?hl=en&lr=&id=Ci0TDgAAQBAJ&oi=fnd&pg=PR5&dq=MetcalfeJ,ShimamuraAP+(1994)+Metacognition:knowingaboutknowing.+Cambridge,+MA:+Massachusetts+Institute+of+Technology.&ots=qFaEbvPrSr&sig=jA7Llno0GwVYLB6BidHca9_5KlQ&redir_esc=y#v=onepage&q&f=false

Metamemory—Psychology—Oxford Bibliographies—Obo. (n.d.). Retrieved December 2, 2019, from https://www.oxfordbibliographies.com/view/document/obo-9780199828340/obo-9780199828340-0238.xml

Method of loci—Wikiwand. (n.d.). Retrieved December 2, 2019, from https://www.wikiwand.com/en/Method_of_loci

Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of the Affective Norms for English Words (ANEW) for Italian. *Behavior Research Methods*, *46*(3), 887–903. https://doi.org/10.3758/s13428-013-0405-3

Nelson, T. O. (1990). Metamemory: A Theoretical Framework and New Findings. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 26, pp. 125–173). https://doi.org/10.1016/S0079-7421(08)60053-5

Pannu, J. K., & Kaszniak, A. W. (2005). Metamemory Experiments in Neurological Populations: A Review. *Neuropsychology Review*, *15*(3), 105–130. https://doi.org/10.1007/s11065-005-7091-6

Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW (Affective Norms for English Words). *Behavior Research Methods*, *39*(3), 600–605. https://doi.org/10.3758/BF03193031

Remembering the Grocery Shopping List: A Study on Metacognitive Biases—Mazzoni—1997—Applied Cognitive Psychology—Wiley Online Library. (n.d.). Retrieved December 19, 2019, from https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1099-0720(199706)11:3%3C253::AID-ACP454%3E3.0.CO;2-0

SAYKRS, B. M. (1973). Analysis of Heart Rate Variability. *Ergonomics*, *16*(1), 17–32. https://doi.org/10.1080/00140137308924479

Scherer, K. (2011). On the Rationality of Emotions: Or, When are Emotions Rational? *Social Science Information*, *50*, 330–350. https://doi.org/10.1177/0539018411411018

Scientific Utopia: II. Restructuring Incentives and Practices to Promote Truth Over Publishability—Brian

A. Nosek, Jeffrey R. Spies, Matt Motyl, 2012. (n.d.). Retrieved December 18, 2019, from https://journals.sagepub.com/doi/full/10.1177/1745691612459058

Seo, M.-G., Goldfarb, B., & Barrett, L. F. (2010). Affect and the Framing Effect within Individuals over Time: Risk Taking in a Dynamic Investment Simulation. *Academy of Management Journal*, *53*(2), 411–431. https://doi.org/10.5465/amj.2010.49389383

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*, 258. https://doi.org/10.3389/fpubh.2017.00258

Sharot, T. (2011). The optimism bias. *Current Biology*, *21*(23), R941–R945. https://doi.org/10.1016/j.cub.2011.10.030

Shepperd, J. A., Pogge, G., & Howell, J. L. (2017). Assessing the consequences of unrealistic optimism: Challenges and recommendations. *Consciousness and Cognition*, *50*, 69–78. https://doi.org/10.1016/j.concog.2016.07.004

Valenza, G., Citi, L., Lanatá, A., Scilingo, E. P., & Barbieri, R. (2014). Revealing Real-Time Emotional Responses: A Personalized Assessment based on Heartbeat Dynamics. *Scientific Reports*, *4*(1), 1–13. https://doi.org/10.1038/srep04998