**[Study Information](https://osf.io/9awtb?view_only=" \l "study-information)**

[**Title**](https://osf.io/9awtb?view_only=#study-information.title)

*Provide the working title of your study. It may be the same title that you submit for publication of your final manuscript, but it is not a requirement.*

Does emotional valence and arousal influence metacognition for memory?

[**Authors**](https://osf.io/9awtb?view_only=#study-information.authors)

Micah Allen, Camile Correa, Sebastian Scott Engen, Nicolas Legrand

[**Description**](https://osf.io/9awtb?view_only=#study-information.description)

*Please give a brief description of your study, including some background, the purpose of the study, or broad research questions. (optional)*

Emotional and physiological arousal biases both perception and memory. For example, evidence from our lab and others suggest that both pre-stimulus and stimulus-evoked arousal can influence both the decision criterion and confidence during perceptual decisions (Allen et al., 2016; de Gee et al., 2017; Delfini & Campos, 1972). In the domain of memory, arousal exerts a well-known effect on the acuity and content of free recall, for example in the “lightbulb” effect in which memory for traumatic, exciting, or otherwise emotionally arousing events is heightened (Heuer & Reisberg, 1990). However, to our knowledge, no experiment has investigated whether arousal also influences metacognition for memory (metamemory). This project will adapt a pre-existing task (McCurdy et al., 2013), for measuring the sensitivity, bias, and efficiency of metamemory for arousing versus unarousing stimuli (Fleming & Lau, 2014). Further, we will test whether stimulus-evoked heart-rate changes mediate the influence of arousal on metacognition.

We will adapt a commonly used measure of meta-memory, see e.g., McCurdy et al., (2013) as a key reference. Participants will make a 2AFC judgment for item pairs, to discriminate which have been seen previously in a memorized list. We will adapt this task so that some blocks involve emotionally neutral trials and others high-arousing (positive and negative valence) words. The order of block will be pseudorandomized and counter-balanced across participants to ensure no block to block repetition of arousal condition. On each trial, target words will be presented together with previously unseen “distractor” words of equivalent valence and arousal.

[**Hypotheses**](https://osf.io/9awtb?view_only=#study-information.hypotheses)

*List specific, concise, and testable hypotheses. Please state if the hypotheses are directional or non-directional. If directional, state the direction. A predicted effect is also appropriate here. If a specific interaction or moderation is important to your research, you can list that as a separate hypothesis.*

Manipulation Checks:

Stimulus Driven Arousal

To check for a stimulus-driven effect on overall arousal (e.g., that high vs low arousal word stimuli increase physiological arousal), we will perform a one-sample t-test (one-sided) comparing the mean effect of word arousal on heart-rate acceleration (see Physiological Preprocessing and Analysis). We will also perform a full 2x2 ANOVA on the Valence by Arousal interaction, to explore possible interactions of stimulus valence on the cardiac response to arousing stimuli.

Hypothesis testing:

Our principal tests of interest will be 2x2 (Arousal by Valence) repeated measures ANOVAs on memory sensitivity (d-prime; d’), average confidence, and M-ratio. D’ and M-ratio 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. Additionally, we will explore fully hierarchical models, implementing the 2x2 ANOVA within the HMM framework.

Null Hypothesis: No Effect of Emotion on Meta-Memory

Under the null hypothesis, we expect to observe no effects 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. 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.

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

Under this hypothesis, we expect to see a main effect of stimulus arousal on 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 we would also expect heart-rate 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 Meta-Memory

Under this hypothesis, we also expect to see a main effect of stimulus arousal on recall accuracy and confidence. However, here we expect this effect to depend 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 posthoc pairwise t-test (e.g., PositiveHighArousal > NegativeHighArousal and PositiveLowArousal vs NegativeHighArousal). Under this account, the sharpening effects of arousal on stimulus representation interact with the biasing influence of stimulus valence. Here we would additionally expect heart-rate to mediate the influence of accuracy on confidence in a valence-dependent interaction.

[**Design Plan**](https://osf.io/9awtb?view_only=#design-plan)

[**Study type**](https://osf.io/9awtb?view_only=#design-plan.study-type)

*Please check one of the following statements*

* Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.

[**Blinding**](https://osf.io/9awtb?view_only=#design-plan.blinding)

*Blinding describes who is aware of the experimental manipulations within a study. Mark all that apply.*

* No blinding is involved in this study.

[**Is there any additional blinding in this study?**](https://osf.io/9awtb?view_only=#design-plan.is-there-any-additional-blinding-in-this-study?)

*Blinding (Other) (optional)*

[**Study design**](https://osf.io/9awtb?view_only=#design-plan.study-design)

The experiment design is within-subject and takes place in two components. In the first component (60 minutes), the metamemory task, participants complete 12 blocks of the cued-recall task with short breaks in between each block.

For the meta-memory task, we will adapt a commonly used measure of meta-memory, see e.g., https://www.jneurosci.org/content/33/5/1897 as a key reference (Baird, Smallwood, Gorgolewski, & Margulies, 2013; McCurdy et al., 2013; Rouault, McWilliams, Allen, & Fleming, 2018). Participants will make a two-alternative forced-choice judgment for item pairs, to discriminate which of the two stimuli have been seen previously in a memorized list. We will adapt this task so that each block will alternate pseudo-randomly between each level of our 2 by 2 factorial design, with the factors: Valence (Positive vs Negative) and Arousal (High vs Low). The block order will be pseudorandomized and counterbalanced across participants to ensure no block to block repetition of arousal condition.

At the beginning of each block of 50 trials, participants study a list of words arranged in 10 rows and 5 columns.

In each trial, participants view two words presented simultaneously to the left and right of fixation; one “target” word from the study list and the other “distractor word” which had not. For each block, distractor and target words are selected from lists of equivalent arousal and valence, with the selection of target and distractor lists counterbalanced across participants. In this way, all words served as either distractor or target, depending on the participant. Additionally, to partially control for difficulty, on each block participants could study the target list for 0.5, 1, or 1.5 minutes. Four different list selection orders and four different study time orders were counterbalanced across participants, to control for order effects.

Subsequently, participants rate how confident they are that their 2-AFC judgment was correct using a 7-point Likert scale. Both responses have to be provided within 3 s. Stimuli will be the top 1200 most and least arousing words, divided equally by positive and negative valence, all derived from the “Affective Norms for English Words” database.

We calculated the upper and lower 32% tertile in both Valence and Arousal and thereby divided the ANEW dictionary into five categories of (See picture attached): High Valence and High Arousal words Low Valence and High Arousal words High Valence and Low Arousal words Low Valence and Low Arousal words Mildly Arousing and Mildly Valenced words The first four categories including high/low Arousing/Valence words were then divided into 6 lists of 50 words, each list having approximately the same median as the other lists within the same category (See picture attached).

This division enables the 12 block design where each category will go through 3 blocks showing 3 lists of 50 studied words and 3 lists of 50 distractor words. There are 300 words in total for each of the four categories.

Because a full-blown double translation and revalidation in Danish are not possible for this study, all participants will be selected for having a high level of English fluency, and measures of heart rate and subjective arousal ratings by stimulus condition will provide manipulation checks.

However, 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). However, as all participants will also provide new ratings for these stimuli, we will repeat the primary analyses (ANOVA on heart-rate, confidence, d’, and meta-memory) after reclassifying all stimuli categories based on the participants' ratings, if they are significantly different from the ANEW normed scores.

These ratings are done as an at-home post-experiment survey, where we measure the participants' ratings of valence and arousal using a web-based version of the original procedure used in the original ANEW survey (Bradley, & Lang, 1999).

The pictures attached show the screens presented to the participants. It includes a 9-point Likert scale, the word being rated as well as the original drawings of SAM all of which were used in the original ANEW survey. Moreover, there’s no time limit for responses and the survey is carried out using Pavlovia.org - the online version of the PythonScript Editor PsychoPy.

*(optional)*

* [1 Task.png](https://osf.io/project/9awtb/files/osfstorage/5d9c3c98a7bc73000be6b630/)
* [2 List.png](https://osf.io/project/9awtb/files/osfstorage/5d9c3c98a7bc73000be6b62e/)
* [3 trial.png](https://osf.io/project/9awtb/files/osfstorage/5d9c3c98a7bc73000be6b62c/)
* [4 Sampling.png](https://osf.io/project/9awtb/files/osfstorage/5d9c3c98a7bc73000be6b62a/)
* [5 table rating.png](https://osf.io/project/9awtb/files/osfstorage/5d9c3c98a7bc73000be6b628/)

[**Randomization**](https://osf.io/9awtb?view_only=#design-plan.randomization)

*If you are doing a randomized study, how will you randomize, and at what level?  (optional)*

All subjects will perform all memory blocks. The order of the blocks will be randomized, but everyone will perform the same trials.

[**Sampling Plan**](https://osf.io/9awtb?view_only=#sampling-plan)

[**Existing Data**](https://osf.io/9awtb?view_only=#sampling-plan.existing-data)

*Preregistration is designed to make clear the distinction between confirmatory tests, specified prior to seeing the data, and exploratory analyses conducted after observing the data. Therefore, creating a research plan in which existing data will be used presents unique challenges. Please select the description that best describes your situation. Please see https://cos.io/prereg for more information.*

* Registration prior to creation of data

[**Explanation of existing data**](https://osf.io/9awtb?view_only=#sampling-plan.explanation-of-existing-data)

*If you indicate that you will be using some data that already exist in this study, please describe the steps you have taken to assure that you are unaware of any patterns or summary statistics in the data. This may include an explanation of how access to the data has been limited, who has observed the data, or how you have avoided observing any analysis of the specific data you will use in your study. (optional)*

[**Data collection procedures**](https://osf.io/9awtb?view_only=#sampling-plan.data-collection-procedures)

Participants (n=35) will be recruited through advertisements via Sona system (CFIN) from Aarhus University and also via social media (Facebook). Participants must be at least 18 years old, give written consent, be normal or corrected to normal vision and fluent in English. Participants will be paid 100 DKK per hour and the estimated total duration of the test session is 1,5 hour (150 DKK). Participants who complete the at-home valence and arousal ratings will receive an additional 50 DKK.

*(optional)*

* No files selected

[**Sample size**](https://osf.io/9awtb?view_only=#sampling-plan.sample-size)

*Describe the sample size of your study. How many units will be analyzed in the study? This could be the number of people, birds, classrooms, plots, interactions, or countries included. If the units are not individuals, then describe the size requirements for each unit. If you are using a clustered or multilevel design, how many units are you collecting at each level of the analysis?*

Our target sample is 30 participants. We will attempt to recruit 35, assuming a rate of 15% data loss due to poor performance or technical issues with recording.

[**Sample size rationale**](https://osf.io/9awtb?view_only=#sampling-plan.sample-size-rationale)

*This could include a power analysis or an arbitrary constraint such as time, money, or personnel. (optional)*

Our sample size is based on practical considerations; as this is a pilot project for a student thesis project, time and financial resources are limited. However, given our exclusively within-subject design N = 30 should deliver sufficient power to detect medium or larger effect sizes. Furthermore, as the project is pre-registered we hope that it can deliver informative estimates of effect size for future studies in this area.

[**Stopping rule**](https://osf.io/9awtb?view_only=#sampling-plan.stopping-rule)

*If your data collection procedures do not give you full control over your exact sample size, specify how you will decide when to terminate your data collection.  (optional)*

We will stop when we finish collecting data from 35 participants, or when we run out of testing room bookings, whichever comes first.

[**Variables**](https://osf.io/9awtb?view_only=#variables)

[**Manipulated variables**](https://osf.io/9awtb?view_only=#variables.manipulated-variables)

*(optional)*

We manipulate the valence and arousal of stimuli by selecting words from the ANEW database with high levels of valence and arousal, per their normed ratings. Additionally, we manipulate arousal and valence by presenting blocks of words sampled from only one stimulus category within a block, and by preventing arousal blocks from repeating one after the other. See Section 8 for more details on stimulus selection.

*(optional)*

* No files selected

[**Measured variables**](https://osf.io/9awtb?view_only=#variables.measured-variables)

With a trial by trial two-alternative forced-choice, we will ask participants which word did they see on the studied list and how confident they were in their decision (on a scale 1 - guess to 7 - certain). They will later rate arousal (on a scale 1 - calm to 9 -excited) and valence (on a scale 1-unhappy to 9 - happy) of the words.

*(optional)*

* No files selected

[**Indices**](https://osf.io/9awtb?view_only=#variables.indices)

*(optional)*

We measure trial by trial confidence ratings, choices (e.g., if word 1 or word 2 is presented), and reaction times. Additionally we measure pulse oximetry data to estimate heart-rate. In at an at-home post-experiment survey, we measure participants ratings of valence and arousal using a web-based version of the original procedure used in the original ANEW survey and other recent validations. (Montefinese, Ambrosini, Fairfield, & Mammarella, 2014; Moors et al., 2013; Bradley, Lang, Bradley, & Lang, 1999; Redondo, Fraga, Padrón, & Comesaña, 2007).

*(optional)*

* No files selected

[**Analysis Plan**](https://osf.io/9awtb?view_only=#analysis-plan)

[**Statistical models**](https://osf.io/9awtb?view_only=#analysis-plan.statistical-models)

Behavioral Analyses

Trials with RTs faster than 100ms will first be removed from analysis. Next, any trial with RTs greater than 3 standard deviations than the median RT value will be removed. Following these trial level exclusions, for each subject, type-1 performance on the 2AFC task will be calculated using signal-detection theoretic measures of sensitivity (d’), criterion/bias (c), as well as reaction time for each trial, separately for each Arousal by Valence condition. These measures collectively summarize the overall accuracy, bias, and speed of memory-judgements in the task. Metacognitive measures of meta-dprime (md’), meta-ratio (mratio), and meta-criterion (mc) will also be calculated summarizing the sensitivity, efficiency, and bias of confidence judgements. All Type-1 and Type-2 SDT measures (d’, c, md’, mratio, and mc) will be derived from the hierarchical metacognition model (HMM) implemented in Python (Fleming, 2017), ran in the non-hierarchical mode to allow frequentist analysis of the resultant parameters. The HMM has been shown to improve estimates of SDT parameters even when run in non-hierarchical mode compared to the MLE model (Maniscalco & Lau, 2012). Finally, we will calculate the average confidence for each trial.

Statistical Analyses

Our primary analyses will consist of 2x2 Repeated Measures ANOVAs on d’, m-ratio, average confidence, and reaction time. These will be conducted in JASP using frequentist statistical tests, alpha level = 0.95. Additionally, we will split trials into correct vs incorrect trials and estimate a 2 (valence) by 2 (arousal) by 2 (correct/error) ANOVA on average confidence ratings. In the case of any null-effects (see Hypothesis, above) we will re-run these tests using the equivalent Bayesian test under the default cauchy prior = 0.707 in order to assess the evidence for the null hypothesis.

Exploratory Behavioral Analyses

To explore the influence of random effects on these estimates, we will additionally fit multilevel linear models in R, testing for overall effects of arousal, valence, and their interaction on confidence and accuracy, while including random slopes and offsets for each subject, i.e., testing for the full factorial interactions between valence, arousal, and decision accuracy on confidence. Significant effects of each main effect and interaction will be calculated by likelihood ratio tests comparing models with and without these effects included.

Physiological Analysis

Preprocessing

Pulse oximetry data will be used to estimate heart rate variability over long period, and heart rate changes over shorter interval. The raw signal will be converted to interbeat intervals using automated peak detection approaches. The data will be systematically manually checked to ensure the correct estimation of the heart beats. At each peak, we will derive the interbeat interval from the previously detected peak. This measure, expressed in milliseconds, is a reliable estimation of the R-R interval usually measured in electrocardiography.

In the first part of the analyses, we will use these interbeat intervals to compute time, frequency domain, and non-linear indexes of the heart rate variability using custom Python script and the Kubios Software (version 2.2). The following indexes will be used to compare the different experimental conditions: RMSSD, pNN50, mean HR, high (0.15-0.4 Hz) and low (0.04-0.15 Hz) frequency power of heart rate variability (peak frequency, normalised and not normalized power), and the SD1 and SD2 indexes of poincaré plot. Some of these indexes are often highly correlated with each other and reflect the overall contribution of either the parasympathetic system or a mixture between parasympathetic and sympathetic influence. However, given their prevalence in the literature and their connection with physiological arousal, here we decided to report all experimental effects over these indexes at the group level.

In the second set of analyses, we will focus on the heart rate fluctuation induced by the task on each trial. Using the interbeat intervals, we will interpolate a continuous measure of the heart rate frequency expressed in beats per minute. This time series will then be epoched from -1 to 5 seconds after the trial onset and will be normalized using the value at T(0) as a baseline (Δ-BPM).

Statistical Analysis

As a manipulation check, we will perform a one-sided paired samples t-test comparing average Δ-BPM between the high and low arousal condition, collapsed for valence condition, across all participants. This will reveal whether the comparison of arousing vs unarousing words resulted in significant heart-rate acceleration during the trial. Next, we will perform repeated measures ANOVA (Factors: Valence & Arousal) on Δ-BPM to explore any possible interaction of the arousing effect with valence. Here the F-test will also allow the possibility of significant deceleration effects.

Exploratory Analysis

To test for the mediating effect of stimulus-evoked heart-rate on metacognition, we will conduct multilevel mediation analyses. Here our primary model will be according to the figure attached. By testing if the a, b, c’, and c parameters are significantly different from zero, this will reveal whether stimulus-evoked changes in heart-rate partially or fully mediated the predictive influence of choice accuracy on subjective confidence. We will also explore whether the inclusion of reaction-time or stimulus valence in the model influences these parameter estimates.

Statistical Reporting

All effect sizes will be summarized using the partial-η squared statistic or Cohens-d statistic (where appropriate), and all tested effects will be reported in terms of mean, 95% confidence interval, and effect size.

*(optional)*

* [Statistics.png](https://osf.io/project/9awtb/files/osfstorage/5d9c3c98a7bc73000be6b626/)

[**Transformations**](https://osf.io/9awtb?view_only=#analysis-plan.transformations)

*If you plan on transforming, centering, recoding the data, or will require a coding scheme for categorical variables, please describe that process. (optional)*

M-ratio is often strongly non-normal; in this case, we would perform a log-transformation for any linear models.

[**Inference criteria**](https://osf.io/9awtb?view_only=#analysis-plan.inference-criteria)

*What criteria will you use to make inferences? Please describe the information you’ll use (e.g. specify the p-values, Bayes factors, specific model fit indices), as well as cut-off criterion, where appropriate. Will you be using one or two tailed tests for each of your analyses? If you are comparing multiple conditions or testing multiple hypotheses, will you account for this?  (optional)*

We will use the p<.05 criteria for determining if the ANOVA and the post hoc tests indicate that the results are significantly different from the null hypothesis of no effect. In some cases we will supplement these analyses with Bayesian ANOVAs and/or t-tests, considering BFs > 3 or < 0.33 as informative.

[**Data exclusion**](https://osf.io/9awtb?view_only=#analysis-plan.data-exclusion)

*How will you determine which data points or samples if any to exclude from your analyses? How will outliers be handled? Will you use any awareness check? (optional)*

Exclusion criteria: Participants with mean d' less than or equal to 0 will be excluded because this indicates that they are unable to do the task. Participants with mean meta-d' less than or equal to 0 will be excluded because m-ratio is uninterpretable in these cases. Participants with extreme type 1 or 2 hit or false alarm rates ( under 0.05 or greater than 0.95) will be excluded because m-ratio becomes unstable in these cases. If m-ratio cannot be estimated, for example because there is no variability in confidence reports or in accuracy, then that participant will be excluded. Participants with missing data (i.e. data loss due to technical failures) or insufficient data (fewer than 50 trials in total per condition) will be excluded. Finally, outliers in task performance will be detected by inspecting boxplots of 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 will be excluded from the analysis.

[**Missing data**](https://osf.io/9awtb?view_only=#analysis-plan.missing-data)

*How will you deal with incomplete or missing data? (optional)*

All included subjects will complete all measures.

[**Exploratory analysis**](https://osf.io/9awtb?view_only=#analysis-plan.exploratory-analysis)

*If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time. (optional)*

Additional Exploratory Analyses

We may additionally fit hierarchical metacognition models in which the full valence by arousal factorial is accounted for with group priors, and/or reduced models incorporating only the arousal within-subject effect.

We may also fit GLMS to heart-rate responses at a single-trial level, correcting for multiple comparisons using a cluster-corrected permutation approach, inclusion p-value > 0.01. This would enable us to explore when in time heart-rate is modulated by our experimental factors, and when these modulations correlate with trialwise confidence.

We may plot a Bayesian Correlation Matrix to explore between-subject correlations in M-ratio between our different task categories, and also together with heart-rate. These analyses would be highly exploratory and at risk of heightened type-II error, and so would only be reported in supplemental results.

[**Other**](https://osf.io/9awtb?view_only=#other)

[**Other**](https://osf.io/9awtb?view_only=#other.other)

*If there is any additional information that you feel needs to be included in your preregistration, please enter it here. Literature cited, disclosures of any related work such as replications or work that uses the same data, or other context that will be helpful for future readers would be appropriate here. (optional)*

References

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>

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

de Gee, J. W., Colizoli, O., Kloosterman, N. A., Knapen, T., Nieuwenhuis, S., & Donner, T. H. (2017). Dynamic modulation of decision biases by brainstem arousal systems. ELife, 6, e23232. <https://doi.org/10.7554/eLife.23232>

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>

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>

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. Memory & Cognition, 18(5), 496–506. <https://doi.org/10.3758/BF03198482>

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Consciousness and Cognition, 21(1), 422–430. <https://doi.org/10.1016/j.concog.2011.09.021>

McCurdy, L. Y., 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>

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>

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