

39 et al., 2013; Ye et al., 2018). Memories are fragile internal signals that are prone to substantive
40 decay over time (Davis & Zhong, 2017; Otgaar et al., 2019). Episodic recall can be biased by
41 emotion and arousal both in the context of encoding (Yonelinas & Ritchey, 2015) and during
42 active recall (Ochsner, 2000). The metacognition of memory recall can, in turn, be influenced
43 by the level of details and the “feeling-of-knowing” associated with an episode (Chua et al.,
44 2014; Reggev et al., 2011). In the context of witness testimony for example, if the suspect had
45 an unremarkable face or the memory was hazy, then a witness may report lower confidence in
46 their recollection. A reliable witness should therefore not only recall events as experienced in
47 detail but also accurately assess the fidelity or confidence associated with those memories. As
48 little is currently known about the ability to self-monitor memory for emotional stimuli, we
49 conducted a confirmatory, pre-registered investigation of emotional metamemory.

50 In controlled laboratory settings, classic metacognition experiments often require
51 participants to view a stimulus, make a decision (e.g., whether the stimulus is known or
52 unknown), and report their confidence in this judgement. Healthy individuals typically display
53 reasonably accurate metacognitive insight and achieve a high correlation between confidence
54 and accuracy, even in the absence of external feedback. Metacognition tasks have been applied
55 to investigate a variety of cognitive domains including visual perception (Allen et al., 2016;
56 Fleming et al., 2015), memory (Baird et al., 2013; Fleming et al., 2014; McCurdy et al., 2013),
57 and value-based decision-making (De Martino et al., 2013). However, even with these simple
58 lab-based tasks, participants exhibit substantive interindividual differences in metacognitive
59 ability, and a variety of manipulations can reliably dissociate confidence and accuracy by
60 biasing subjective confidence reports (Fleming et al., 2015; Rollwage et al., 2020).

61 Though little is known about how emotion influences metacognition, previous
62 investigations of memory and emotion highlight stimulus valence and arousal as likely sources
63 of bias. For example, words rated as highly arousing or positively valenced are recognized and
64 detected faster (Kever et al., 2019), and the emotional content of valenced words, either positive
65 or negative, can increase the learner’s confidence in subsequent accurate recall (Tauber &
66 Dunlosky, 2012). Similarly, physiological arousal can promote the awareness of concepts
67 congruent with the bodily state (Kever et al., 2015, 2017), and the level of arousal at encoding
68 can enhance subsequent memory performance (Anderson, 2005; Cahill & McGaugh, 1998) via
69 greater amygdala activity (Kensinger & Corkin, 2004). Accordingly, flashbulb memories (i.e.,
70 vivid and detailed memories encoded under arousing conditions) are recalled more easily and
71 with less decay under specific circumstances (Shields et al., 2017; Yonelinas & Ritchey, 2015).
72 This line of evidence suggests that emotional content, especially those of a highly arousing or
73 negative nature, could bias the salience of the memory signal during recall. This can ultimately
74 result in overconfidence which, in the context of testimony, could bias the individual when
75 estimating the accuracy of his or her recall.

76 Additionally, a core aspect of emotion is that it often coincides with and is triggered by
77 changes in internal bodily states (James, 1884), which is expressed by indices of autonomic
78 activity such as cardiac or respiratory frequency (Kreibig, 2010; Valenza et al., 2014). Heart
79 rate, for example, can be altered both when perceiving emotional stimuli and during their
80 encoding and recollection (Abercrombie et al., 2008; Critchley et al., 2005; Legrand et al.,
81 2020). These bodily changes exert a substantial effect on the mapping between confidence and
82 decision accuracy, which can ultimately also bias metacognition. For example, both
83 experimental and pharmacological modulations of arousal have been shown to bias

84 metacognitive insight, modulating confidence for error trails in a visual task (Allen et al., 2016;
85 Hauser et al., 2017).

86 Here, we hypothesized that both the valence and arousal of an encoded stimulus might
87 modulate the accuracy of memory itself, and investigated whether healthy individuals are
88 aware of such emotional effects on their recognition accuracy. To test this hypothesis, we
89 conducted a pre-registered experiment in which participants memorised lists of words grouped
90 by their levels of valence and arousal. Although most metamemory research has relied on
91 “feeling of knowing” self-report measures, these can be subject to substantive biases, i.e. such
92 as conflating self-report bias with metacognitive sensitivity or being confounded by overall
93 accuracy level (Fleming & Lau, 2014). To overcome these issues, we adapted a signal-theoretic
94 modelling approach to estimate recognition metamemory for emotional versus unemotional
95 words. If arousal primarily biases memory, for example by increasing the salience of encoded
96 items, we would expect to observe a positive main effect of item arousal on both accuracy and
97 metacognitive confidence. Conversely, if emotion primarily biased metacognition through a
98 valence-specific ‘anchoring’ effect, we would expect to observe a full interaction of stimulus
99 arousal and valence on both measures. As a third alternative, if metacognition were robust to
100 emotional biases, we would expect to observe the effects of stimulus valence and arousal on
101 accuracy and response speed, but not on confidence or metacognition. To complement these
102 analyses, we further recorded cardiac measures of physiological arousal through pulse
103 oximetry, to assess their mediating effect on the association between confidence and accuracy.
104

105 **Materials and methods**

106 *Pre-registration and Open Materials*

107 To improve our control of type-I and type-II error rates, as well as the overall reproducibility
108 of the study, the trial was pre-registered before any data collection using the standard Open
109 Science Foundation template. Detailed information regarding power analysis, sample size
110 considerations, experimental and trial design, planned analyses and other key points can be
111 found at the following URL: (<https://osf.io/9awtb>). In what follows, Confirmatory Analyses
112 and Results refer to planned analyses detailed in the pre-registration, whereas Exploratory
113 Analyses and Results refer to post-hoc exploratory analyses conducted following contact with
114 the data. Additionally, in the case of any minor deviation from the pre-registration, these are
115 documented on a case by case basis.

116 *Participants*

117 Thirty-five participants (26 females) between the ages of 18 and 26 ($M = 21$, $SD = 1.9$) were
118 recruited through local advertisements and took part in the experiment at Aarhus University
119 Hospital, Denmark. From the total sample of 35 participants, a sub-set of 30 participants passed
120 the pre-registered exclusion criteria and were analyzed further. All participants had normal or
121 corrected to normal vision, were fluent in English and provided informed written consent
122 before the experiment. The procedures were conducted following the Declaration of Helsinki
123 and with approval from the Danish Neuroscience Centre’s (DNC) Institutional Review Board

124 (IRB). Participants received monetary compensation of 100 DKK per hour. The estimated total
125 duration of the test session was 1,5 hours (150 DKK). Participants also completed a post-test
126 stimulus validation measure in which they provided valence and arousal ratings for all stimuli
127 for an additional 50 DKK. All 35 participants completed the follow-up rating experiment.

128

129 ***Procedure***

130 The experimental procedure included one laboratory and one at-home survey session on two
131 different days with one week in between. In the laboratory session, participants completed a
132 word recognition metamemory task designed to assess the effects of valence and arousal on
133 verbal recognition memory and metacognition. In the survey session, participants rated their
134 subjective feelings of valence and arousal evoked by the words used during the laboratory
135 session.

136 At the beginning of the laboratory session, participants were briefed on the nature of the
137 investigation, were provided task instructions and completed a brief training session of the
138 metamemory task. The training included an example learning phase of 50 neutral and
139 unarousing words, followed by an example testing phase of 10 trials with confidence ratings
140 (see Metamemory Task and Stimuli).

141 During the metamemory task, heart rate was monitored using a Nonin 3012LP Xpod USB
142 pulse oximeter together with a Nonin 8000SM 'soft-clip' fingertip sensor
143 (<https://www.nonin.com/>) attached to the left index finger.

144 ***Word selection***

145 Stimuli consisted of 1200 English words selected from the Affective Norms for English Words
146 (ANEW) database based on valence and arousal ratings measured among a population of
147 American students (Bradley & Lang, 1999). Although ANEW is not validated in the Danish
148 population, previous standardization of the database in Dutch, Spanish and Italian populations
149 (Montefinese et al., 2014; Moors et al., 2013; Redondo et al., 2007) showed good consistency
150 across both American and European samples. In order to estimate the reliability of these ratings
151 among our sample, we conducted an at-home valence and arousal rating task after the
152 behavioural test. We created 4 distinct subgroups of 300 word stimuli, according to a 2 by 2
153 factorial design, where the factors corresponded to valence (positive vs. negative) and arousal
154 (low vs. high). To this aim, we used the tertile of the valence and arousal distribution to exclude
155 words with intermediate ratings, whose valence and arousal might be ambiguous (see Fig. 2a).

156 ***Metamemory Task***

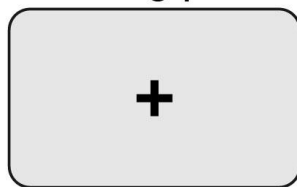
157 Participants completed a word recognition metamemory task adapted from a previous study
158 (McCurdy et al., 2013) to test the influence of emotional valence and arousal on memory and
159 metacognition. The task included 12 blocks, each consisting of a learning phase (Fig. 1a) and
160 a testing phase (Fig. 1b). In the learning phase, participants viewed a list of 50 English words
161 for durations of 30, 60 or 90 seconds. The words were presented on the screen in the form of a
162 table containing five columns with ten rows of words each, and the participants were instructed
163 to memorize as many words as possible. The list of words in each learning phase corresponded
164 to a unique combination of the following factorial conditions: Valence (positive, negative),
165 Arousal (low, high). Participants were notified when 10 seconds of learning time was left by

166 the display of a small warning at the bottom of the screen. During the testing phase, participants
 167 completed 50 trials designed to measure recognition memory and metamemory. On each trial,
 168 two-word stimuli were presented to the left and right of a fixation cross. The word pair
 169 consisted of a “target” and a “distractor”, corresponding to words that were present or absent
 170 in the previous learning phase, respectively. Target and distractor words were matched by
 171 valence and arousal, and their position was randomized across trials. Participants were
 172 instructed to press either the left or the right arrow key to indicate which of the two words they
 173 recognized from the memorised list. This procedure corresponds to a two-alternative forced-
 174 choice task (2AFC) design, which provides optimal conditions for estimating and comparing
 175 metacognition scores across tasks (Lee et al., 2018). Following the button press, participants
 176 provided a subjective confidence rating from 1 (“not confident at all/guessing”) to 7 (“very
 177 confident”). Both button presses and confidence ratings had a maximum time-limit of 3s. If
 178 participants had slower responses, a brief message (i.e., “too slow!”) was displayed on the
 179 screen and the trial was marked as missed.

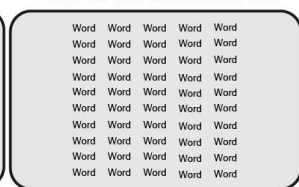
180 The blocks were presented in a pseudo-randomized order to ensure that high arousal blocks
 181 were systematically interleaved with low-arousal blocks. The block order and the selection of
 182 target vs. distraction lists were counterbalanced across participants.
 183

A. Experimental design

Learning phase

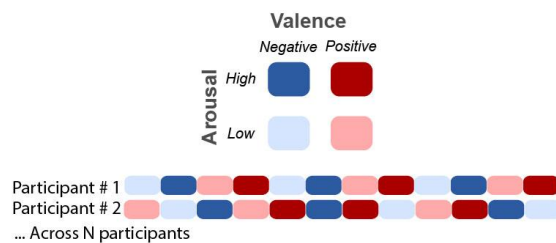


List to be memorized

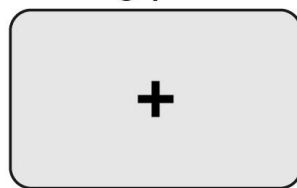


Learning times: 30/60/90 s

Counterbalanced conditions blocks

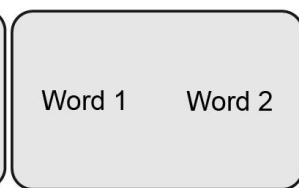


Testing phase



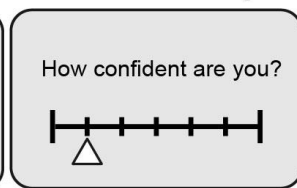
0.25 s

Old/new word decision

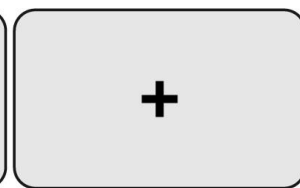


1.5-3.0 s

Confidence rating



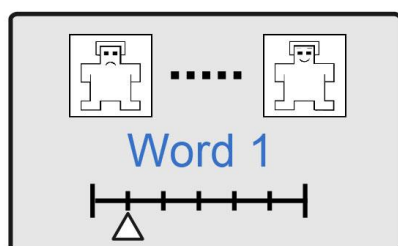
1.5-3.0 s



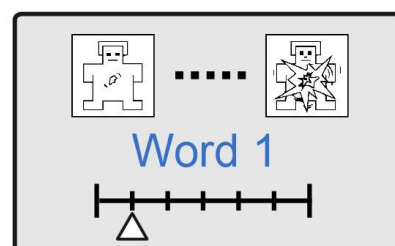
1.75-2.25 s

B. Post-experimental survey

Valence rating



Arousal rating



184

185

186 **Figure 1: A.** Experimental design. The metamemory task contained 12 experimental blocks, each consisting of a
187 learning phase and a testing phase for the 50 words, in a factorial design separated by each Valence and Arousal
188 condition. To limit habituation effects, block orders were counter-balanced in a pseudo-randomized order such
189 that each high arousal block was interceded by a low arousal condition. **B.** Post-experimental survey. To validate
190 our stimulus categories with respect to the original ANEW ratings, participants completed a short subject visual
191 analogue scale rating of valence and arousal for the 1200 words used in the main task (600 target and 600
192 distractors) in an at-home experiment. This was done in a web-based version of the original procedure used in the
193 original ANEW survey.

194 *Valence and Arousal Rating Task*

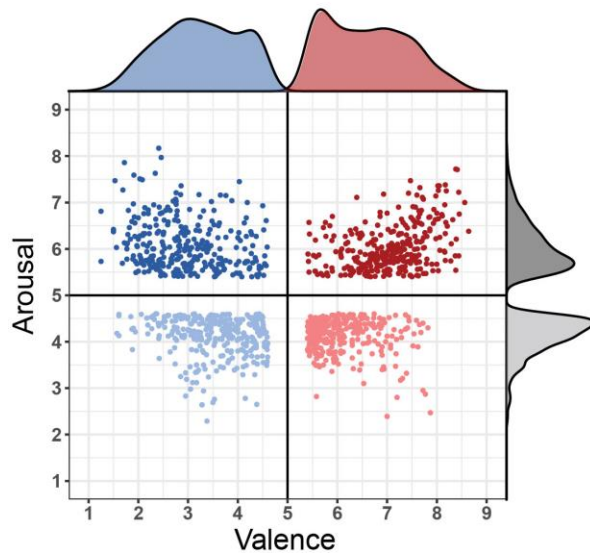
195 To validate the arousal and valence stimulus categories in our Danish sample, participants
196 completed an at-home valence and arousal subjective rating task. They were instructed to
197 provide valence and arousal ratings of their subjective experience associated with each word
198 presented in the metamemory task. The ratings were collected using a 9-point visual numerical
199 scale in a web-based version of the original ANEW survey protocol (Bradley & Lang, 1999).
200 Our version was implemented using Pavlovia (<https://pavlovia.org>), an online platform for
201 running PsychoPy experiments (Peirce et al., 2019). Each word was presented twice, once for
202 valence and once for arousal, and the 9-point scales were complemented with pictures of the
203 original drawings of the Self-Assessment Manikin (Bradley & Lang, 1994), as in the original
204 ANEW survey (Bradley & Lang, 1999). Participants rated a total of 1200 words, self-pacing
205 through all rating trials. We compared the ratings provided by the participants in this study with
206 the normative ANEW ratings using a Spearman rank correlation test (see Fig. 2 c & d). After
207 inspecting histograms of participant responses, we excluded one participant, who only ever
208 pressed the same key (rating 5); this exclusion criterion was not noted in the pre-registered
209 protocol. Overall, stimulus ratings in our sample corresponded very well to the original ANEW
210 ratings, $\rho = [0.93-0.78]$, albeit with lower overall consistency for the arousal vs. valence
211 dimension (Fig. 2).

212

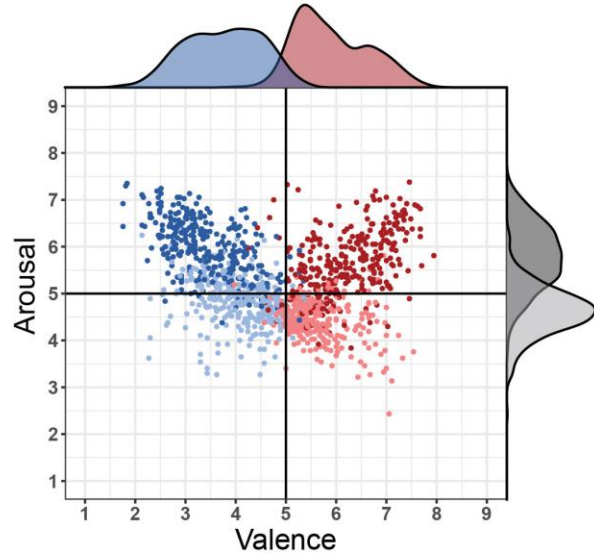
213

214

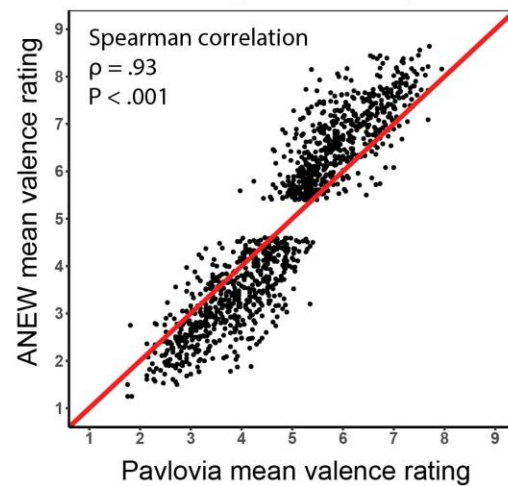
A. ANEW word ratings



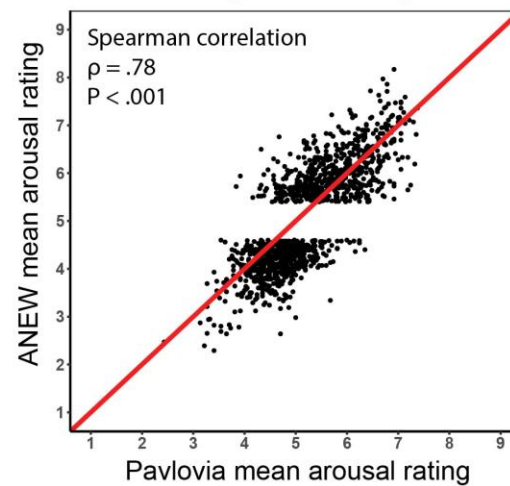
B. Pavlovia word ratings



C. Valence rating consistency



D. Arousal rating consistency



215

216

217 **Figure 2: Stimulus selection and rating validation.** Two rating procedures were used to select and validate word

218 stimuli; the original ANEW normed ratings and the PAVLOVIA at-home ratings completed by our participants in

219 the post-experimental survey. Each word was rated on a 9-point scale (1-9) for valence and arousal separately. **A**

220 **& B.** We selected the words used in the metamemory task by removing items from the central tertile in the arousal and valence rating distributions (panel A). The blue and red dots represent words with negative and positive

221 valence, respectively. The light and dark points represent low and high arousal, respectively. The densities

222 represent the distribution for positive and negative valence (red and blue), and arousal (light and dark). **C & D.**

223 We compared the independent rating provided by the ANEW database to the actual ratings provided by the

224 participants, PAVLOVIA, after the main procedure. Both ratings of valence and arousal showed reasonably high

225 consistency, $\rho = [0.93-0.78]$. See online article for colour figures. The black dots represent each word in the

226 datasets and the red line shows the identity line.

227 **Signal Theoretic Metacognition Modelling**

228 Here we applied a signal-theoretic computational model to describe participant behaviour on

229 the metamemory task (Fleming, 2017; Maniscalco & Lau, 2012). This approach delineates

230 overall behaviour into 'type-I' and 'type-II' measures, corresponding to a basic discrimination

231 performance versus metacognitive levels of performance (Galvin et al., 2003). The type-I

232 performance was quantified using reaction times (RTs) and the signal-detection theoretic

233 (SDT) measures of d-prime (D') and criterion (c) (Macmillan & Creelman, 2004). Type-II

234 performance (i.e., metacognition) was assessed by the SDT measures of Meta-d' and Meta-
235 ratio (M-ratio) (Fleming & Lau, 2014). Briefly, this approach models metacognitive “hits”
236 (e.g., high/low confidence for correct/error trials, respectively) and “misses” (e.g., high/low
237 confidence for error/correct trials, respectively). All SDT-based measures were estimated at
238 the subject level (Fleming, 2017). This model has been extensively described and validated
239 previously (Fleming, 2017; Mazancieux et al., 2020; Morales et al., 2018), here we recount the
240 approach in the context of the present study to aid interpretation.

241 D-prime or d' is a measure of a participant's sensitivity to detect previously studied words
242 during the learning phase, independently from subjective response biases. Instead, criterion or
243 c' encodes the participant's response bias, that is, the overall tendency to prefer one response
244 over the other (e.g. if a participant chose the word presented to the left of the fixation point
245 more often than the alternative). Together with measures of reaction time, d' and c' are metrics
246 of “first-order” or “type-I” task performance. In contrast, Meta-d is an estimate of the
247 sensitivity of subjective confidence ratings to type-I performance (i.e., the probability to be
248 highly confident when correct, or uncertain when incorrect). Meta-d' is, therefore, a measure
249 of insight, or how well one can consciously discriminate their own type-I performance (Lau &
250 Rosenthal, 2011). However, metacognitive sensitivity is also a function of the overall
251 perceptual signal, and as such is substantively influenced by differences in d'. To control for
252 this effect, the ‘M-ratio’ (Meta-d' divided by d') is estimated as a measure of metacognitive
253 efficiency, denoting how a subject's metacognitive sensitivity over- or underperforms what can
254 be expected given their type-I sensitivity (Fleming & Lau, 2014). Finally, average confidence
255 on each condition denotes participants “meta-criterion” or “meta-c”, or their overall level of
256 metacognitive bias denoting the tendency to be confident or uncertain irrespective of accuracy.
257 Meta d' and meta-c are metrics of “second-order” or “type-II” performance.

258

259 *Confirmatory Analyses*

260 Metamemory task

261 All data were pre-processed according to the protocols established in our pre-registration.
262 Accordingly, we excluded all trials with reaction times (RT) faster than 100 ms, greater than 3
263 standard deviations from the median RT, and missing data (absence of response or because the
264 response button was pressed too early or too late). Due to an unforeseen technical error, an
265 absence of response in some trials contaminated the following trial, resulting in negative
266 response times. These trials were also automatically rejected. This procedure resulted in the
267 exclusion of 3.49% (± 3.68) of the trials. Finally, outliers in task performance for each of the
268 conditions were detected based on reaction time, d', and confidence distributions. We also
269 excluded participants showing any extreme value using Tukey's boxplots. Based on these
270 criteria, 5 participants were excluded from all behavioural analyses. These preprocessing steps
271 are also extensively described in the interactive Jupyter notebooks made available on the OSF
272 repository: <https://osf.io/9awtb>.

273 The preprocessing of the behavioural data was carried out using custom R scripts, using R
274 Studio (1.2.5019), the R software (R 3.6.1), and Python scripts using Python 3.7.6. The
275 Bayesian and frequentist statistical models were implemented using the JASP software

276 (<https://jasp-stats.org/>) version 0.12.2 and the R package (AFEX 0.27-2). All Type-1 and Type-
277 2 SDT measures (d' , criterion, meta- d' , m-ratio, and mc) were derived from the hierarchical
278 meta-cognition model (HMM) (Fleming, 2017) implemented in R
279 (<https://github.com/metacoglab/HMeta-d>), run on the individual level to enable frequentist
280 analysis of the resultant parameters.

281 Heart Rate Monitoring

282 We monitored instantaneous heart rate variability using a Nonin 3012LP Xpod USB pulse
283 oximeter together with a Nonin 8000SM 'soft-clip' fingertip sensor (<https://www.nonin.com/>).
284 Pulse oximeters indirectly measure peripheral blood oxygen saturation. The abrupt cyclic
285 increase of oxygenation reflects blood pulse following cardiac contraction. Here, we used the
286 pulse-to-pulse intervals to estimate the instantaneous heart rate. Oxygenation saturation level
287 was continuously recorded at a 75 Hz sampling rate. The preprocessing of the pulse oximetry
288 recording was carried out using Python scripts (Python version 3.7.6) and version 0.1.1 of the
289 Systole Python package (Legrand & Allen, 2020). Statistical analyses were carried out using
290 the Pingouin Python package (Vallat, 2018) and MNE Python (Gramfort, 2013). PPG signals
291 were first upsampled to 1000 Hz and clipping artefacts were corrected using spline
292 interpolation following recent recommendations (van Gent et al., 2019). The signal was then
293 squared for peak enhancement and normalized using the mean + standard deviation using a
294 rolling window (window size: 0.75 seconds). All positive peaks were labelled as systolic
295 (minimum distance: 0.2 seconds). We then detected ectopic, long, short, missed and extra beats
296 using adaptive thresholds over the successive beats-to-beats interval (Lipponen & Tarvainen,
297 2019), as implemented in Systole (Legrand & Allen, 2020). The code implementing these steps
298 can be found in the Jupyter notebooks made available on the OSF repository:
299 <https://osf.io/9awtb>.

300 **Instantaneous Pulse Rate.** All pulses labelled as missed or extra beats were corrected by
301 adding or removing beats, respectively. We then interpolated the instantaneous heart rate at 75
302 Hz to a continuous recording using the previous values and divided it into epochs (from -1
303 second pre-trial to 6 seconds after the word presentation). All the epochs that contained, or
304 were adjacent to, an interval that was labelled as long, short or (pseudo-)ectopic beats were
305 automatically rejected, resulting in an average rejection rate of 18.22% ($\pm 11.49\%$). The
306 instantaneous heart rate was then averaged across trials for each condition and downsampled
307 to 5 Hz for subsequent analyses.

308 **Linear regression.** In an exploratory analysis, we used the instantaneous pulse rate as a
309 predictor of confidence over time to track the relationship between cardiac frequency
310 modulation and metamemory. We extracted the data following the same procedure, this time
311 using 1s before the trial start as a baseline and using the initial sampling rate (75 Hz) to facilitate
312 cluster-based statistical tests. Cluster-based permutation testing was performed using the
313 *permutation_cluster_test()* and the *permutation_cluster_1samp_test()* functions from the MNE
314 Python package (Gramfort, 2013). This enabled us to assess significant point-to-point
315 deviations from zero in encoded responses while controlling for multiple comparisons.

316 **Pulse Rate Variability.** Besides the analysis of the instantaneous pulse rate, we also
317 performed pulse rate variability analyses. Although targeting a different physiological signal
318 as compared to a classic electrocardiogram (ECG), the varying length of pulse cycle provides

319 a sufficiently accurate estimation of the underlying heart rate variability (HRV) when used at
320 rest for healthy young participants (Schäfer & Vagedes, 2013). Here, we extracted the systolic
321 peak intervals using the method presented above. Intervals labelled as missed or extra beats
322 were corrected by adding or removing beats, respectively. Additionally, intervals that were
323 labelled as short, long, or (pseudo-)ectopic beats were corrected using linear interpolation.
324 Following our specification in the pre-registration, we reported heart rate variability metrics in
325 the time (RMSSD, pnn50) and frequency domain (normalized and non-normalized high and
326 low-frequency power), as well as non-linear indexes (SD1 and SD2). These indexes reflect
327 changes in beat-to-beat intervals and measure sympathetic and parasympathetic influences on
328 the heart (Shaffer et al., 2014; Shaffer & Ginsberg, 2017). We inspected the resulting time
329 series and rejected noisy and unreliable segments (2 segments were rejected in total). The
330 values of each metric were then averaged across learning time (30, 60 or 90 seconds), and the
331 summary variables were entered into a 2 by 2 repeated-measures ANOVA with factors stimulus
332 arousal (arousing vs. unarousing) and valence (positive vs. negative).

333 **Results**

334 *Behavioural Results*

335 Overview

336 Following our pre-registration, the behavioural analyses focused on two levels of performance
337 during the metamemory task: type-I variables corresponding to the discrimination ability, and
338 type-II variables describing metacognition. To assess memory performance, we analyzed
339 decision accuracy, discrimination sensitivity (d'), bias (c), and response time (RT). To assess
340 metacognition, we analyzed average confidence (i.e., metacognitive bias), metacognitive
341 sensitivity (Meta- d'), and metacognitive efficiency (M-ratio, Meta- d'/d'). All signal theoretic
342 measures (d' , c , Meta- d' , and M-ratio) were estimated using a unified Bayesian approach, as
343 described previously (Fleming, 2017). All posthoc tests were corrected for multiple
344 comparisons using the Holm procedure. Here, we reported only the key details of the significant
345 effects; full ANOVA tables and associated statistics for all analyses can be found in our JASP
346 notebooks located online at the following URL: <https://osf.io/pefnr/>.

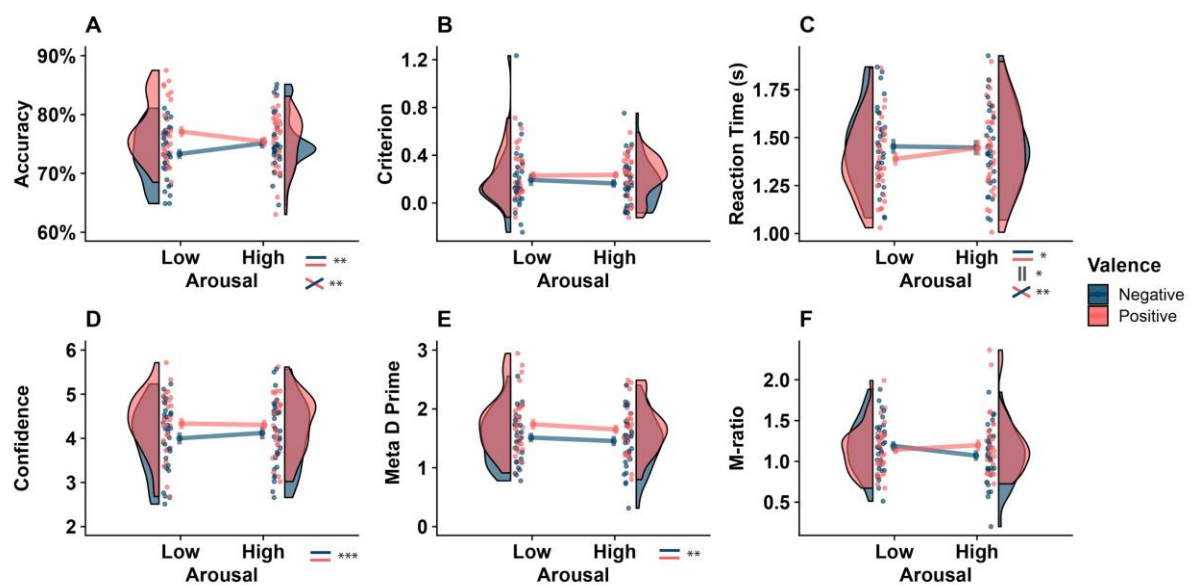
347 Recognition Memory (Type-I)

348 First, we examined the influence of emotional Valence and Arousal on decision accuracy, in a
349 two-way repeated measures ANOVA, collapsing across block and learning time conditions.
350 We found a significant main effect of Valence ($F_{(1,29)} = 9.887$, $\eta_p^2 = 0.254$, $p = 0.004$), as well
351 as a significant interaction between Valence and Arousal ($F_{(1,29)} = 7.779$, $\eta_p^2 = 0.212$, $p = .009$),
352 as positive words were recognized more accurately than negative ones under low arousal
353 conditions ($T_{(29)} = 4.20$, $p_{\text{Holm}} < 0.001$). Similarly, for sensitivity (d'), we found a significant
354 effect of Valence ($F_{(1,29)} = 11.34$, $\eta_p^2 = 0.28$, $p = 0.002$), as negative words decreased d' , as well
355 as an interaction between Valence and Arousal ($F_{(1,29)} = 7.34$, $\eta_p^2 = 0.20$, $p = 0.011$), as positive
356 words were recognized more sensitively than negative ones under low arousal conditions
357 ($T_{(29)} = 4.304$, $p_{\text{Holm}} < 0.001$). When analyzing the median response time, we observed a
358 significant effect of Valence ($F_{(1,29)} = 0.55$, $\eta_p^2 = 0.16$, $p = .025$), an effect of Arousal ($F_{(1,29)} =$

359 6.94, $\eta_p^2 = 0.19$, $p = .013$) and an interaction between these two factors ($F_{(1,29)} = 7.56$, $\eta_p^2 =$
360 0.20, $p = .010$), revealing that participants responded faster to positive valence under the low
361 compared to the high arousal condition ($T_{(29)} = 3.80$, $p_{\text{Holm}} = .002$). Analysis of response criterion
362 revealed no significant main effects or interactions, and no other significant effects were found
363 (all $p_s > .05$).
364

365 Metacognition (Type-II)

366 We then performed a second level of analysis on the metacognition data, comprising average
367 confidence, meta-d and M-ratio. First, we performed a Valence \times Arousal repeated measures
368 ANOVA on the average confidence. This procedure showed a strong effect of Valence ($F_{(1,29)}$
369 $= 14.98$, $\eta_p^2 = 0.34$, $p < 0.001$), as participants were more confident for positive valenced words.
370 No other effects or interactions were significant. Participants were also more sensitive to their
371 performance (Meta-d') when responding to positive valenced words (main effect of Valence
372 $F_{(1,29)} = 11.28$, $\eta_p^2 = 0.28$, $p = 0.002$). Concerning the M-ratio (i.e., Meta-d'/d'), which measures
373 metacognitive efficiency, we found no main effect or interactions (all $p_s > 0.05$). Following
374 our pre-registered protocol, we followed up this analysis with a Bayesian ANOVA (Rouder et
375 al., 2012) implemented in JASP (version 0.12.2) (JASP Team, 2020), to assess the strength of
376 evidence for the null effect. This analysis compares the evidence for nested models of
377 increasing complexity; e.g., comparing a null model to those with only main effects of valence
378 or arousal, or a full model with main effects and interaction terms. This revealed strong relative
379 evidence for the null overall model (including subject offsets), $\text{BF}_{\text{Model}} = 7.28$; the next best
380 model was one with a main effect of Valence whose relative $\text{BF}_{\text{Model}} = 0.74$, i.e. inconclusive
381 evidence. This analysis suggests that under the default JASP priors, it is very unlikely that
382 Valence, Arousal, or their interaction exerted any effect on metacognitive efficiency.
383
384



385
386 **Fig. 3.** Behavioural results showing factorial main effects and interactions on discrimination and metacognitive
387 performance. Modified raincloud plots (Allen et al., 2019) illustrating behavioural results of discrimination
388 measures of accuracy (A), criterion (B) and reaction time (C) as well as metacognitive measures of confidence
389 (D), Meta-d' (E) and M-ratio (F). Repeated measures ANOVA (Valence \times Arousal) was carried out for each

390 condition separately. The upper panel shows that a significant main effect of emotional valence was observed as
391 negative valenced words reduced accuracy (A) and slowed down reaction times (C). Similarly, the lower panel
392 shows a main effect of valence for both Confidence and Meta-d' are impaired by negative valence. (** $p < 0.001$,
393 ** $p < 0.01$, * $p < 0.05$).

394 Encoding duration

395 The effects of the encoding phase duration on both type I and type II task performances were
396 not included in the pre-registration and so were assessed in an exploratory post hoc analysis.
397 The results are described in the Supplementary Materials.

398

399 *Physiological results*

400 Pulse rate variability

401 First, we analyzed the effect of valence and arousal on the heart rate frequency during the
402 experimental blocks. We averaged the estimated beats per minute (Mean BPM) across the
403 different learning times (30, 60 and 90 seconds) and submitted it to a two-way repeated measure
404 ANOVA (Valence \times Arousal). Results showed a main effect of Valence ($F_{(1,29)} = 10.852$, $\eta_p^2 =$
405 0.272 , $p = 0.003$), meaning lower BPM for negative valenced words, but no other main or
406 interaction effects.

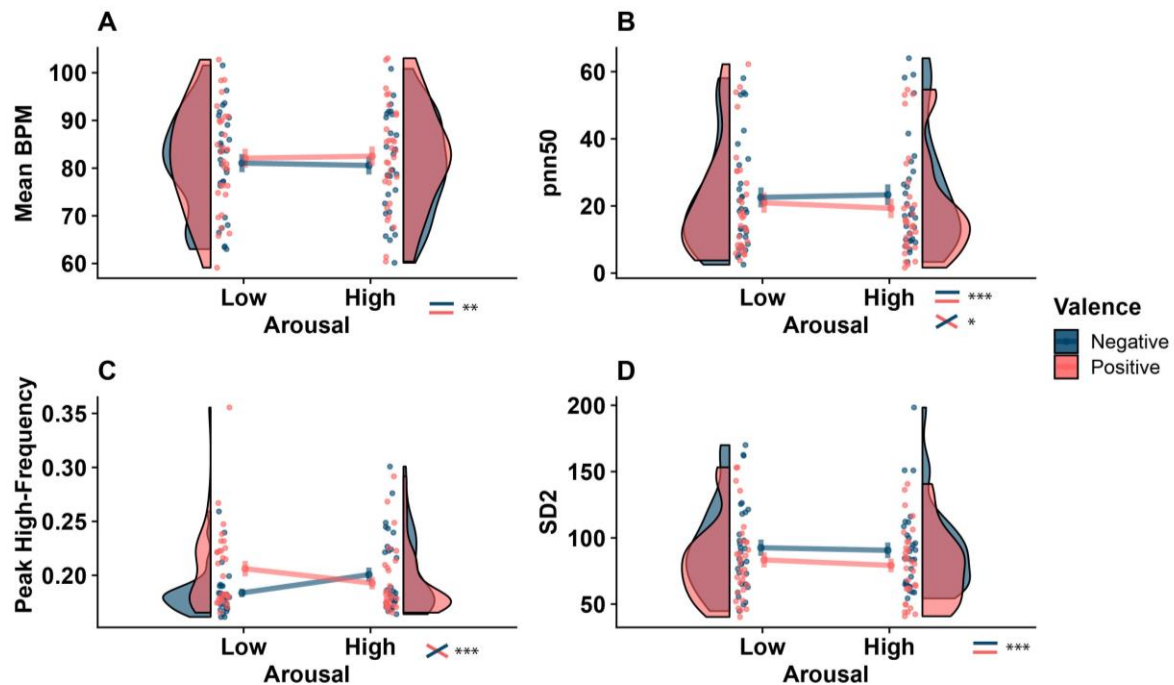
407 For the low and high-frequency peak analysis, the peak high-frequency (HF peak) revealed
408 an interaction between valence and arousal ($F_{(1,29)} = 14.50$, $\eta_p^2 = 0.33$, $p < 0.001$) such that high-
409 frequency cardiac oscillations were suppressed by negative emotional valence under low but
410 not high arousal ($T_{(29)} = 3.36$, $p_{\text{Holm}} = .007$).

411 Concerning the Root Mean Square of the Successive Differences (RMSSD, not shown in
412 Fig. 4), we found a main effect of valence ($F_{(1,29)} = 6.74$, $\eta_p^2 = 0.19$, $p = .015$) i.e. negative
413 valence increased RMSSD, but no other main effect or interaction (all $ps < .05$).

414 When considering the proportion of successive beat-to-beat intervals deviating by more than
415 50 ms (pnn50) we observed an effect of Valence ($F_{(1,29)} = 24.17$, $\eta_p^2 = 0.45$, $p < 0.001$), as well
416 as an interaction between Valence and Arousal ($F_{(1,29)} = 4.54$, $\eta_p^2 = 0.13$, $p = .042$). Under high
417 arousal the positive valence suppressed pnn50 while negative valence increased it ($t_{(29)} = -4.98$,
418 $p_{\text{Holm}} < 0.001$).

419 Finally, we also analyzed the effect of these factors on the non-linear metrics of heart rate
420 variability SD1 and SD2. The SD2 metric revealed an effect of Valence ($F_{(1,29)} = 35.20$, $\eta_p^2 =$
421 0.55 , $p < 0.001$), so that negative valence increased SD2 heart rate variability, but we found no
422 other main effects or interactions. Concerning SD1, we found no significant effects (all $ps >$
423 $.05$). These results are illustrated in **Fig. 4**; here we reported the main significant effects,
424 however full analyses details and results tables can be found in the HRV JASP notebook
425 located on the Github repository for this study.

426



427
428
429
430
431
432
433
434
435
436
437

Fig. 4: Modified raincloud plots illustrating results of pulse rate variability (PRV) analyses. PRV indices were calculated separately for each 50 trial block and averaged by condition. Mean BPM (A), Pnn50 (B), High-frequency peak (C), SD2 (D). Repeated measures ANOVA (Valence \times Arousal) was then carried out for each variable separately. A significant main effect of emotional valence was observed for mean BPM, as negative valence decreased cardiac activity frequency, as well as for the pnn50 and the non-linear SD2 metric. We did not observe a main effect of Arousal, but an interaction with valence was found for the high-frequency peak, such that high-frequency cardiac oscillations were reduced by negative emotional valence under low but not high arousal. No other significant effects were found. (***) $p < 0.001$, ** $p < 0.01$, * $p < 0.05$). See *Methods* and *PRV Results* for more details.

438 Event-related analysis

439 Next, we analyzed the time-locked instantaneous pulse rate fluctuation following word
440 presentation. **Figure 5a** shows the evoked cardiac frequency fluctuation following the display
441 of the two words on the screen. Following the specification of the pre-registered report, we
442 analyzed the average of this fluctuation across time (**Figure 5b**). Here, we observed no effect
443 of Valence, ($F_{(1,28)} = 0.441$, $\eta_p^2 = 0.015$, $p = 0.511$), Arousal ($F_{(1,28)} = 0.003$, $\eta_p^2 < 0.001$, $p =$
444 0.954) or an interaction between these two factors ($F_{(1,28)} = 0.044$, $\eta_p^2 = 0.001$, $p = .833$).

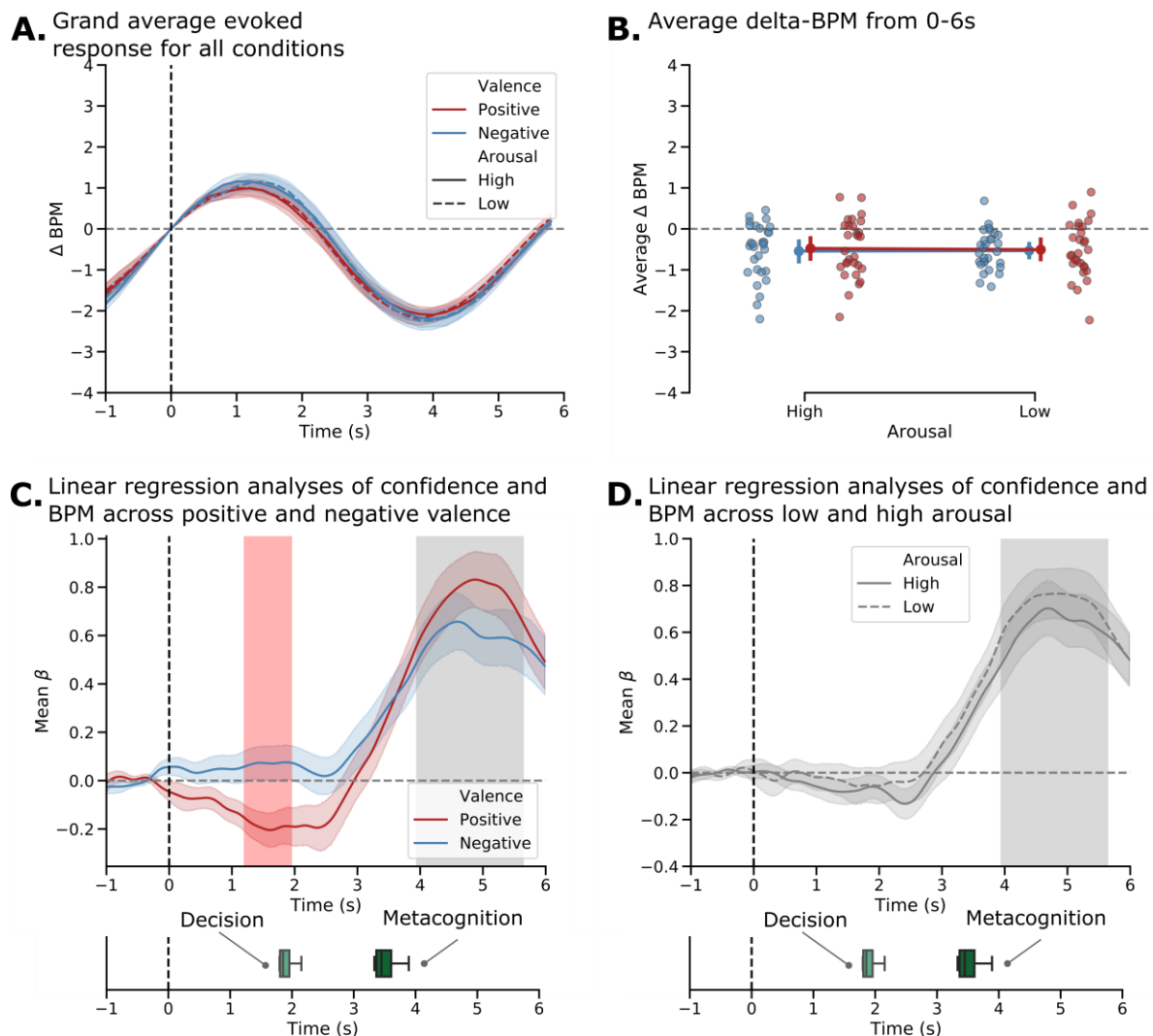
445 Linear regression

446 In an additional exploratory analysis, we also tested the possible interaction between the
447 instantaneous pulse rate modulation observed during decision and metacognition and the
448 subjective report provided by the participant. For each participant and condition separately, we
449 used the reported confidence C and the instantaneous pulse rate BPM at each time point s of
450 the trial t to fit a linear regression of the form:

451
452
453

$$\text{BPM}_{s,t} = \alpha_s + \beta_s \times C_t + \varepsilon$$

454 All variables were normalized and beta-values for each explanatory regressor and participant
 455 were extracted for statistical analysis. First, to test for a difference in beta values from 0 across
 456 time across all conditions, i.e. the average effect of trial by trial confidence reports on
 457 fluctuations in evoked heart-rate, replicating previous analysis linking these variables (Allen et
 458 al., 2016). We averaged between conditions assessed significance via non-parametric cluster-
 459 level t-test. Results show a significant cluster (3.94-5.65 seconds after stimulus presentation, p
 460 = 0.001). As our HRV and behavioural results emphasized an effect of stimulus valence on
 461 both metacognitive behaviour and cardiac activity, we then compared the association between
 462 confidence and instantaneous cardiac activity between different valence and arousal conditions.
 463 When comparing positive and negative valence conditions (averaging across arousal levels)
 464 we found a significant early cluster (1.20-2.96 seconds after stimulus presentation, $p=0.047$),
 465 suggesting that stimulus valence modulates the correlation between evoked heart-rate and
 466 confidence. Finally, we repeated this comparison for high vs. low arousal conditions,
 467 collapsing stimulus valence. This analysis found no significant clusters. See Fig. 5 for
 468 illustration of these results.



469 **Fig. 5: Modulation of the cardiac activity at the trial level and its relation with reported subjective**
 470 **confidence.** A. Evoked pulse rate activity shows that the overall experimental procedure modulated the
 471 instantaneous cardiac frequency over time through an early acceleration component after the trial start (0-2 s) and
 472 a later deceleration component (2-6 s). This pattern is consistent with an orientation reflex, suggesting early brief
 473 integration and later sensory or memory processing. Interestingly, these two components are also time-locked
 474 with the decision and metacognition average response time. Here, we did not observe any difference between the
 475

476 experimental conditions. **B.** We averaged the instantaneous pulse rate in the window of interest (0-5 s) and
477 confirmed this absence of effect and an overall diminution of cardiac frequency after the trial start. **C.** Beta values
478 over time of the linear regression (Confidence ~ BPM) for positive and negative valence trials separately. The
479 confidence level was associated with the instantaneous cardiac frequency during the late time window
480 corresponding to the metacognition decision. **D.** Beta values over time of the linear regression (Confidence ~
481 BPM) for high and low arousal trials separately. Using the same approach, contrasting for High and Low level of
482 arousal. Significance assessed using a cluster-level statistical permutation test ($\alpha=0.05$). Shaded areas and
483 error bar show the 68% CI. Significant clusters are shown by a shaded red path for condition contrast, and grey
484 path for null tests. See online article for colour figures.

485 Discussion

486 In this study, we investigated the influence of emotion on word memory and metamemory
487 through a combination of experimental psychology, cognitive modelling, and physiology. We
488 adapted a recognition memory paradigm (McCurdy et al., 2013) where participants memorised
489 lists of words varying in arousal and valence. Stimulus valence exerted a consistent influence
490 on recognition performance, metacognitive confidence, and physiological activity. For reaction
491 time and accuracy, this effect was greatest for low vs. high arousal words, suggesting that the
492 influence of stimulus valence on memory depends in part on arousal. On the physiological
493 level, we observed an association between the subjective confidence reported by the participant
494 and the evoked pulse rate, which was also modulated by the word valence. Our results evidence
495 that although recognition memory is impaired for negative emotional stimuli, participants can
496 accurately monitor and report this uncertainty. Further, monitoring the effect of emotion on
497 memory may depend in part on integrating the associated changes in cardiac signals.

498 The notion that emotional events are better recalled than neutral ones has been extensively
499 discussed (Yonelinas & Ritchey, 2015), but the psychophysiological mechanisms mediating
500 this effect are still unclear. In some cases, emotional influences on memory have been assumed
501 to be almost entirely ascribed to arousal, which is presumed to control the cognitive and
502 behavioural importance of stimuli (Mather, 2007; Mather & Sutherland, 2009), while other
503 works attributed this improvement primarily to valence (Adelman & Estes, 2013; Kensinger,
504 2009). Previous investigations in the perceptual domain have documented that arousing stimuli
505 “boost” the signal-to-noise ratio of visual motion, as reflected in both models of ballistic
506 evidence accumulation, and subjective confidence (Allen et al., 2016; Lufityanto et al., 2016).
507 However, in our study, the effect of arousal was generally muted or dependent on stimulus
508 valence.

509 Whereas on type-I task performance, stimulus valence interacted with arousal, for
510 metacognitive type-II variables we observed only a pronounced main effect of valence with no
511 arousal effect or interaction. In general, participant confidence reports closely matched the
512 overall effect of stimulus emotion on performance; negative valence decreased sensitivity,
513 increased reaction times, and decreased confidence. The robust evidence we observed for the
514 absence of effect on metacognitive-efficiency (M-ratio) further underlines this finding; the
515 strong null Bayes factor here demonstrates that shifts in subjective confidence were well
516 reflected by the magnitude of any changes in type-I sensitivity, indicating that subjects make
517 optimal use of the available memory signal during metacognitive judgements, irrespective of
518 any conditional valence or an arousal effect. This finding suggests that, although memory is

519 degraded under negative emotional contexts, participants can accurately account for this in
520 their subjective confidence.

521 One possible explanation for the absence of arousal effect is found in our validation rating
522 study (see Figure 2); while the valence dimension was well preserved between the ANEW
523 database and the ratings by our participants after the task, the consistency of arousal ratings
524 was slightly reduced. This limits the extent to which we can infer actual stimulus arousal in our
525 data, and it may be that the stimuli were simply not sufficiently distinct or triggering for a
526 Danish sample. Indeed, in this study arousing versus nonarousing word stimuli did not evoke
527 a significant difference in physiological arousal response. Future studies could benefit from
528 both a larger corpus of validated words, a more general sample of English-speaking
529 participants, and multiple modalities of memorised stimuli which may better preserve arousal-
530 based dimensions. Here, it should also be noted that the use of words instead of images or film
531 is also a potential confounding factor due to the lack of vividness and complexity of the mental
532 imagery elicited by one single item. While previous works evidenced that words classified as
533 highly arousing and with extreme positive or negative valence are associated with better
534 memory recall (Buchanan et al., 2006; Madan et al., 2017, 2019), other dimensions like
535 semantic properties or the functional use of the object such as animacy are more relevant
536 psycholinguistic dimensions to predict free recall (Madan, 2020). Again, controlling for this
537 dimension in future studies could help to refine the influence of emotion on metamemory
538 beyond the dimensions of valence and arousal.

539 In a related line of research, several investigations have linked physiological activity (e.g.,
540 as indexed by pupil dilation or cardiac acceleration) to subjective confidence and
541 metacognition. According to influential predictive-coding accounts of metacognition (Allen et
542 al., 2016; Meyniel et al., 2015; Moulin & Souchay, 2015), confidence reflects the width of a
543 posterior decision variable, such that fluctuations in arousal bias the gain or precision of this
544 distribution. Here, we examined both trial level evoked changes in cardiovascular activity and
545 summary measures of pulse rate variability separately for each condition. When examining
546 instantaneous heart rate variation, we observed a robust sinusoidal pattern that remained stable
547 across conditions, similar to an orientation reflex triggered by trial onset (see Figure 5). The
548 shape and intensity of this cardiac deceleration can reflect several cognitive processes like
549 attention orienting, emotion processing or inhibitory control (Abercrombie et al., 2008;
550 Critchley et al., 2005; Legrand et al., 2020). Replicating previous findings (Allen et al., 2016),
551 we observed a robust association between trial-by-trial fluctuations in subjective confidence
552 during this late interval, with the strength of this association being modulated by stimulus
553 valence during the early, decision-evoked period. These results suggest that at least some
554 variance in the monitoring of emotional inputs on metamemory could arise from monitoring
555 the associated physiological changes. This result corroborates the notion that memory retrieval
556 is an embodied process (Garfinkel et al., 2013), which has implications both for their
557 conservation, the accuracy of their recall, but also their control in the case of distressing
558 emotional memories (Gagnepain et al., 2017; Legrand et al., 2020).

559 Whereas no overall modulation of instantaneous heart-rate was seen for stimulus valence or
560 arousal, here we observed substantive, robust modulations of heart rate variability (HRV) when
561 subjects recalled negatively valenced stimuli across multiple time, frequency and non-linear
562 indices. HRV (i.e the amount of change across time of the interbeat intervals) can reflect the
563 influence of higher cognitive processes on cardiac frequency through the parasympathetic

564 nervous system (Smith et al., 2017; Thayer & Lane, 2009). Across the different range of HRV
565 indices we examined, two showed a strong valence main effect (i.e., Mean BPM & SD2),
566 whereas others (i.e., high-frequency peak and pnn50) showed a robust interaction between
567 these factors. Although disentangling what underlies these different effects is far from trivial,
568 it is interesting to note the dissociation between these effects, and similarity to those observed
569 for our type-I and type-II metamemory measures. One intriguing possibility is that the high-
570 frequency variability indexed by the former two measures may be a more direct input for
571 metacognitive monitoring than the others, as these showed a similar pattern of exaggerated
572 valence effect with no effect of arousal. One means to probe this hypothesis is to correlate
573 individual differences in the modulation of confidence by valence with each HRV metric;
574 however, our study is underpowered for individual differences analyses (Schönbrodt &
575 Perugini, 2013), leaving it as an intriguing avenue for future research.

576 Several important limitations should be considered when interpreting our HRV effects. As
577 HRV is here calculated by collapsing across each 50 trial block, the modulations observed
578 therein are necessarily a mixture of multiple cognitive states and perceptual inputs; future
579 studies could benefit from disentangling the encoding, perceptual, and retrieval stages to better
580 account for these stages of the decision process. Additionally, here we assessed heart rate
581 variability through pulse oximetry recording. Pulse oximetry recordings are used as an
582 alternative to the electrocardiogram (ECG) by several clinical and non-clinical studies
583 (Quintana, Elstad, et al., 2016). The sampling rate of our device (75 Hz) is not optimal when
584 compared to recommended standards for electrocardiogram (ECG) recording and HRV
585 measurement (Quintana, Alvares, et al., 2016), which could limit our ability to detect true
586 effects, particularly in the lower frequency range. Previous reports, however, have shown a
587 strong consistency between the estimated pulse rate variability and the heart rate variability as
588 measured through ECG (Lu et al., 2009; Schäfer & Vagedes, 2013). Similarly, we did not
589 measure or control respiratory cycles during this study, which robustly modulate HRV
590 measures, in particular in the lower frequency. Collectively, while our results nicely
591 demonstrate that stimulus emotional content modulates high-frequency indices of
592 cardiovascular arousal, future studies in this area are likely to benefit from a combination of
593 more nuanced experimental design and a more sophisticated recording set-up.

594

595 **Conclusion**

596 This pre-registered study sheds light on the biasing effects of valence on metamemory with
597 possible physiological correlates of these effects. Negatively valenced stimuli globally
598 decreased both memory performance and metacognition, supporting a role for emotions in
599 guiding confidence and memory performance. While arousal has often been described as a
600 possible mechanism of the beneficial effect of emotion on memory, we only found limited
601 evidence for performance improvement under highly arousing conditions. In line with these
602 main effects, we found that stimulus valence modulated the overall pulse rate variability and
603 the association between instantaneous heart-rate and subjective confidence. Collectively, these
604 results suggest that although negative stimuli do exert a degrading influence on recognition
605 memory, participants are largely able to account for this effect in their subjective confidence,
606 perhaps by monitoring physiological states.

607
608
609
610
611
612
613
614

Acknowledgements

NL, CMCC, NKM, NN, and MA are supported by a Lundbeckfonden Fellowship (under Grant [R272-2017-4345]), and the AIAS-COFUND II fellowship programme that is supported by the Marie Skłodowska-Curie actions under the European Union’s Horizon 2020 (under Grant [754513]), and the Aarhus University Research Foundation.

615 **ORCIDS:**

- 616 N. L. : 0000-0002-2651-4929
617 S. S. E. : 0000-0002-5348-6417
618 C. M. C. C. : 0000-0001-7663-5828
619 N. K. M. : 0000-0001-6580-5289
620
621 N. N. : 0000-0003-1226-1139
622
623 F. F. : 0000-0002-9974-6261
624 M. A. : 0000-0001-9399-4179

625
626 **Disclosure statement.** No financial interest or benefit that has arisen from the direct
627 applications of this research.

628 **Data availability statement.**

629 The project pre-registration can be found at the following link: <https://osf.io/9awtb>

630 **Data deposition and supplemental online material.**

631 All behavioural and physiological data can be found at: <https://osf.io/pefnr/>
632

633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659

References

- Abercrombie, H., Chambers, A., Greischar, L., & Monticelli, R. (2008). Orienting, emotion, and memory: Phasic and tonic variation in heart rate predicts memory for emotional pictures in men. *Neurobiol Learn Mem*, *90*, 644–650.
- Adelman, J. S., & Estes, Z. (2013). Emotion and memory: A recognition advantage for positive and negative words independent of arousal. *Cognition*, *129*(3), 530–535. <https://doi.org/10.1016/j.cognition.2013.08.014>
- 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*. <https://doi.org/10.7554/elife.18103>
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. <https://doi.org/10.12688/wellcomeopenres.15191.1>
- Anderson, A. K. (2005). Affective Influences on the Attentional Dynamics Supporting Awareness. *Journal of Experimental Psychology: General*, *134*(2), 258–281. <https://doi.org/10.1037/0096-3445.134.2.258>
- Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and Lateral Networks in Anterior Prefrontal Cortex Support Metacognitive Ability for Memory and Perception. *Journal of Neuroscience*, *33*(42), 16657–16665. <https://doi.org/10.1523/JNEUROSCI.0786-13.2013>
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49–59. [https://doi.org/10.1016/0005-7916\(94\)90063-9](https://doi.org/10.1016/0005-7916(94)90063-9)
- Bradley, M. M., & Lang, P. J. (1999). *Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings*.

- 660 Buchanan, T. W., Etzel, J. A., Adolphs, R., & Tranel, D. (2006). The influence of autonomic
661 arousal and semantic relatedness on memory for emotional words. *International*
662 *Journal of Psychophysiology*, 61(1), 26–33.
663 <https://doi.org/10.1016/j.ijpsycho.2005.10.022>
- 664 Cahill, L., & McGaugh, J. L. (1998). Mechanisms of emotional arousal and lasting declarative
665 memory. *Trends in Neurosciences*, 21(7), 294–299. [https://doi.org/10.1016/S0166-](https://doi.org/10.1016/S0166-2236(97)01214-9)
666 [2236\(97\)01214-9](https://doi.org/10.1016/S0166-2236(97)01214-9)
- 667 Chua, E. F., Pergolizzi, D., & Weintraub, R. R. (2014). The Cognitive Neuroscience of
668 Metamemory Monitoring: Understanding Metamemory Processes, Subjective Levels
669 Expressed, and Metacognitive Accuracy. In S. M. Fleming & C. D. Frith (Eds.), *The*
670 *Cognitive Neuroscience of Metacognition* (pp. 267–291). Springer.
671 https://doi.org/10.1007/978-3-642-45190-4_12
- 672 Critchley, H. D., Rotshtein, P., Nagai, Y., O’Doherty, J., Mathias, C. J., &
673 Dolan, R. J. (2005). Activity in the human brain predicting differential heart rate
674 responses to emotional facial expressions. *NeuroImage*, 24(3), 751–762.
675 <https://doi.org/10.1016/j.neuroimage.2004.10.013>
- 676 Davis, R. L., & Zhong, Y. (2017). The Biology of Forgetting—A Perspective. *Neuron*, 95(3),
677 490–503. <https://doi.org/10.1016/j.neuron.2017.05.039>
- 678 De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based
679 choice. *Nature Neuroscience*, 16(1), 105–110. <https://doi.org/10.1038/nn.3279>
- 680 Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency
681 from confidence ratings. *Neuroscience of Consciousness*, 2017(1).
682 <https://doi.org/10.1093/nc/nix007>
- 683 Fleming, S. M., Dolan, R. J., & Frith, C. D. (2012). Metacognition: Computation, biology and
684 function. *Philosophical Transactions of the Royal Society B: Biological Sciences*,
685 367(1594), 1280–1286. <https://doi.org/10.1098/rstb.2012.0021>

- 686 Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human*
687 *Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00443>
- 688 Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-Specific
689 Disruption of Perceptual Confidence. *Psychological Science*, 26(1), 89–98.
690 <https://doi.org/10.1177/0956797614557697>
- 691 Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific
692 impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*,
693 137(10), 2811–2822. <https://doi.org/10.1093/brain/awu221>
- 694 Gagnepain, P., Hulbert, J., & Anderson, M. C. (2017). Parallel Regulation of Memory and
695 Emotion Supports the Suppression of Intrusive Memories. *The Journal of*
696 *Neuroscience*, 37(27), 6423–6441. <https://doi.org/10.1523/JNEUROSCI.2732-16.2017>
- 697 Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal
698 detectability: Discrimination between correct and incorrect decisions. *Psychonomic*
699 *Bulletin & Review*, 10(4), 843–876. <https://doi.org/10.3758/BF03196546>
- 700 Garfinkel, S. N., Barrett, A. B., Minati, L., Dolan, R. J., Seth, A. K., & Critchley, H. D. (2013).
701 What the heart forgets: Cardiac timing influences memory for words and is modulated
702 by metacognition and interoceptive sensitivity: Cardiac timing and memory.
703 *Psychophysiology*, 50(6), 505–512. <https://doi.org/10.1111/psyp.12039>
- 704 Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in*
705 *Neuroscience*, 7. <https://doi.org/10.3389/fnins.2013.00267>
- 706 Hauser, T. U., Allen, M., Purg, N., Moutoussis, M., Rees, G., & Dolan, R. J. (2017).
707 Noradrenaline blockade specifically enhances metacognitive performance. *eLife*, 6,
708 e24901. <https://doi.org/10.7554/eLife.24901>
- 709 Heyes, C., Bang, D., Shea, N., Frith, C. D., & Fleming, S. M. (2020). Knowing Ourselves
710 Together: The Cultural Origins of Metacognition. *Trends in Cognitive Sciences*, 24(5),
711 349–362. <https://doi.org/10.1016/j.tics.2020.02.007>

- 712 James, W. (1884). What is an emotion? *Mind*, 9(34), 188–205.
713 <https://doi.org/10.1093/mind/os-ix.34.188>
- 714 JASP Team. (2020). *JASP (Version 0.12.2)[Computer software]*. <https://jasp-stats.org/>
- 715 Kensinger, E. A. (2009). Remembering the Details: Effects of Emotion. *Emotion Review*, 1(2),
716 99–113. <https://doi.org/10.1177/1754073908100432>
- 717 Kensinger, E. A., & Corkin, S. (2004). Two routes to emotional memory: Distinct neural
718 processes for valence and arousal. *Proceedings of the National Academy of Sciences*,
719 101(9), 3310–3315. <https://doi.org/10.1073/pnas.0306408101>
- 720 Kever, A., Grynberg, D., Eeckhout, C., Mermillod, M., Fantini, C., & Vermeulen, N. (2015).
721 The body language: The spontaneous influence of congruent bodily arousal on the
722 awareness of emotional words. *Journal of Experimental Psychology: Human*
723 *Perception and Performance*, 41(3), 582–589. <https://doi.org/10.1037/xhp0000055>
- 724 Kever, A., Grynberg, D., Szmalec, A., Smalle, E., & Vermeulen, N. (2019). “Passion” versus
725 “patience”: The effects of valence and arousal on constructive word recognition.
726 *Cognition and Emotion*, 33(6), 1302–1309.
727 <https://doi.org/10.1080/02699931.2018.1561419>
- 728 Kever, A., Grynberg, D., & Vermeulen, N. (2017). Congruent bodily arousal promotes the
729 constructive recognition of emotional words. *Consciousness and Cognition*, 53, 81–88.
730 <https://doi.org/10.1016/j.concog.2017.06.007>
- 731 Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological*
732 *Psychology*, 84(3), 394–421. <https://doi.org/10.1016/j.biopsycho.2010.03.010>
- 733 Lau, H., & Rosenthal, D. (2011). Empirical support for higher-order theories of conscious
734 awareness. *Trends in Cognitive Sciences*, 15(8), 365–373.
735 <https://doi.org/10.1016/j.tics.2011.05.009>

- 736 Lee, A. L. F., Ruby, E., Giles, N., & Lau, H. (2018). Cross-Domain Association in
737 Metacognitive Efficiency Depends on First-Order Task Types. *Frontiers in*
738 *Psychology*, 9, 2464. <https://doi.org/10.3389/fpsyg.2018.02464>
- 739 Legrand, N., & Allen, M. (2020). *Systole: V 0.1.2—September 2020* (0.1.2) [Computer
740 software]. Zenodo. <https://doi.org/10.5281/ZENODO.3607912>
- 741 Legrand, N., Etard, O., Vandeveld, A., Pierre, M., Viader, F., Clochon, P., Doidy, F.,
742 Peschanski, D., Eustache, F., & Gagnepain, P. (2020). Long-term modulation of cardiac
743 activity induced by inhibitory control over emotional memories. *Scientific Reports*,
744 *10*(1), 15008. <https://doi.org/10.1038/s41598-020-71858-2>
- 745 Lipponen, J. A., & Tarvainen, M. P. (2019). A robust algorithm for heart rate variability time
746 series artefact correction using novel beat classification. *Journal of Medical*
747 *Engineering & Technology*, 43(3), 173–181.
748 <https://doi.org/10.1080/03091902.2019.1640306>
- 749 Lu, G., Yang, F., Taylor, J. A., & Stein, J. F. (2009). A comparison of photoplethysmography
750 and ECG recording to analyse heart rate variability in healthy subjects. *Journal of*
751 *Medical Engineering & Technology*, 33(8), 634–641.
752 <https://doi.org/10.3109/03091900903150998>
- 753 Lufityanto, G., Donkin, C., & Pearson, J. (2016). Measuring Intuition: Nonconscious
754 Emotional Information Boosts Decision Accuracy and Confidence. *Psychological*
755 *Science*, 27(5), 622–634. <https://doi.org/10.1177/0956797616629403>
- 756 Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide* (2nd ed.).
757 Psychology Press. <https://doi.org/10.4324/9781410611147>
- 758 Madan, C. R. (2020). Exploring word memorability: How well do different word properties
759 explain item free-recall probability? *Psychonomic Bulletin & Review*.
760 <https://doi.org/10.3758/s13423-020-01820-w>

- 761 Madan, C. R., Scott, S. M. E., & Kensinger, E. A. (2019). Positive emotion enhances
762 association-memory. *Emotion*, *19*(4), 733–740. <https://doi.org/10.1037/emo0000465>
- 763 Madan, C. R., Shafer, A. T., Chan, M., & Singhal, A. (2017). Shock and awe: Distinct effects
764 of taboo words on lexical decision and free recall. *Quarterly Journal of Experimental*
765 *Psychology (2006)*, *70*(4), 793–810. <https://doi.org/10.1080/17470218.2016.1167925>
- 766 Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating
767 metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*,
768 *21*(1), 422–430. <https://doi.org/10.1016/j.concog.2011.09.021>
- 769 Mather, M. (2007). Emotional Arousal and Memory Binding: An Object-Based Framework.
770 *Perspectives on Psychological Science*, *2*(1), 33–52. <https://doi.org/10.1111/j.1745->
771 [6916.2007.00028.x](https://doi.org/10.1111/j.1745-6916.2007.00028.x)
- 772 Mather, M., & Sutherland, M. (2009). Disentangling the Effects of Arousal and Valence on
773 Memory for Intrinsic Details. *Emotion Review*, *1*(2), 118–119.
774 <https://doi.org/10.1177/1754073908100435>
- 775 Mazancieux, A., Fleming, S. M., Souchay, C., & Moulin, C. J. A. (2020). Is there a G factor
776 for metacognition? Correlations in retrospective metacognitive sensitivity across tasks.
777 *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000746>
- 778 McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013).
779 Anatomical Coupling between Distinct Metacognitive Systems for Memory and Visual
780 Perception. *Journal of Neuroscience*, *33*(5), 1897–1906.
781 <https://doi.org/10.1523/JNEUROSCI.1890-12.2013>
- 782 Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian Probability: From
783 Neural Origins to Behavior. *Neuron*, *88*(1), 78–92.
784 <https://doi.org/10.1016/j.neuron.2015.09.039>

- 785 Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). The adaptation of
786 the Affective Norms for English Words (ANEW) for Italian. *Behavior Research*
787 *Methods*, 46(3), 887–903. <https://doi.org/10.3758/s13428-013-0405-3>
- 788 Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., van Schie, K., Van Harmelen, A.-L.,
789 De Schryver, M., De Winne, J., & Brysbaert, M. (2013). Norms of valence, arousal,
790 dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods*,
791 45(1), 169–177. <https://doi.org/10.3758/s13428-012-0243-8>
- 792 Morales, J., Lau, H., & Fleming, S. M. (2018). Domain-General and Domain-Specific Patterns
793 of Activity Supporting Metacognition in Human Prefrontal Cortex. *The Journal of*
794 *Neuroscience*, 38(14), 3534–3546. <https://doi.org/10.1523/JNEUROSCI.2360-17.2018>
- 795 Moulin, C., & Souchay, C. (2015). An active inference and epistemic value view of
796 metacognition. *Cognitive Neuroscience*, 6(4), 221–222.
797 <https://doi.org/10.1080/17588928.2015.1051015>
- 798 Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The
799 experience and process of recognizing feelings past. *Journal of Experimental*
800 *Psychology: General*, 129(2), 242–261. <https://doi.org/10.1037/0096-3445.129.2.242>
- 801 Otgaar, H., Howe, M. L., Patihis, L., Merckelbach, H., Lynn, S. J., Lilienfeld, S. O., & Loftus,
802 E. F. (2019). The Return of the Repressed: The Persistent and Problematic Claims of
803 Long-Forgotten Trauma. *Perspectives on Psychological Science*, 14(6), 1072–1095.
804 <https://doi.org/10.1177/1745691619862306>
- 805 Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E.,
806 & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior*
807 *Research Methods*, 51(1), 195–203. <https://doi.org/10.3758/s13428-018-01193-y>
- 808 Quintana, D. S., Alvares, G. A., & Heathers, J. A. J. (2016). Guidelines for Reporting Articles
809 on Psychiatry and Heart rate variability (GRAPH): Recommendations to advance

- 810 research communication. *Translational Psychiatry*, 6(5), e803–e803.
811 <https://doi.org/10.1038/tp.2016.73>
- 812 Quintana, D. S., Elstad, M., Kaufmann, T., Brandt, C. L., Haatveit, B., Haram, M., Nerhus, M.,
813 Westlye, L. T., & Andreassen, O. A. (2016). Resting-state high-frequency heart rate
814 variability is related to respiratory frequency in individuals with severe mental illness
815 but not healthy controls. *Scientific Reports*, 6(1), 37212.
816 <https://doi.org/10.1038/srep37212>
- 817 Redondo, J., Fraga, I., Padrón, I., & Comesaña, M. (2007). The Spanish adaptation of ANEW
818 (Affective Norms for English Words). *Behavior Research Methods*, 39(3), 600–605.
819 <https://doi.org/10.3758/BF03193031>
- 820 Reggev, N., Zuckerman, M., & Maril, A. (2011). Are all judgments created equal?: An fMRI
821 study of semantic and episodic metamemory predictions. *Neuropsychologia*, 49(5),
822 1332–1342. <https://doi.org/10.1016/j.neuropsychologia.2011.01.013>
- 823 Rollwage, M., Loosen, A., Hauser, T. U., Moran, R., Dolan, R. J., & Fleming, S. M. (2020).
824 Confidence drives a neural confirmation bias. *Nature Communications*, 11(1), 2634.
825 <https://doi.org/10.1038/s41467-020-16278-6>
- 826 Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors
827 for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374.
828 <https://doi.org/10.1016/j.jmp.2012.08.001>
- 829 Schäfer, A., & Vagedes, J. (2013). How accurate is pulse rate variability as an estimate of heart
830 rate variability? *International Journal of Cardiology*, 166(1), 15–29.
831 <https://doi.org/10.1016/j.ijcard.2012.03.119>
- 832 Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize?
833 *Journal of Research in Personality*, 47(5), 609–612.
834 <https://doi.org/10.1016/j.jrp.2013.05.009>

- 835 Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and
836 Norms. *Frontiers in Public Health*, 5, 258. <https://doi.org/10.3389/fpubh.2017.00258>
- 837 Shaffer, F., McCraty, R., & Zerr, C. L. (2014). A healthy heart is not a metronome: An
838 integrative review of the heart's anatomy and heart rate variability. *Frontiers in*
839 *Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01040>
- 840 Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal
841 cognitive control and metacognition. *Trends in Cognitive Sciences*, 18(4), 186–193.
842 <https://doi.org/10.1016/j.tics.2014.01.006>
- 843 Shields, G. S., Sazma, M. A., McCullough, A. M., & Yonelinas, A. P. (2017). The effects of
844 acute stress on episodic memory: A meta-analysis and integrative review.
845 *Psychological Bulletin*, 143(6), 636–675. <https://doi.org/10.1037/bul0000100>
- 846 Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of
847 neurovisceral integration. *Neuroscience & Biobehavioral Reviews*, 75, 274–296.
848 <https://doi.org/10.1016/j.neubiorev.2017.02.003>
- 849 Tauber, S. K., & Dunlosky, J. (2012). Can older adults accurately judge their learning of
850 emotional information? *Psychology and Aging*, 27(4), 924–933.
851 <https://doi.org/10.1037/a0028447>
- 852 Thayer, J. F., & Lane, R. D. (2009). Claude Bernard and the heart-brain connection: Further
853 elaboration of a model of neurovisceral integration. *Neuroscience and Biobehavioral*
854 *Reviews*, 33(2), 81–88. <https://doi.org/10.1016/j.neubiorev.2008.08.004>
- 855 Valenza, G., Citi, L., Lanatá, A., Scilingo, E. P., & Barbieri, R. (2014). Revealing Real-Time
856 Emotional Responses: A Personalized Assessment based on Heartbeat Dynamics.
857 *Scientific Reports*, 4(1). <https://doi.org/10.1038/srep04998>
- 858 Vallat, R. (2018). Pinguin: Statistics in Python. *Journal of Open Source Software*, 3(31),
859 1026. <https://doi.org/10.21105/joss.01026>

- 860 van Gent, P., Farah, H., van Nes, N., & van Arem, B. (2019). Analysing Noisy Driver
861 Physiology Real-Time Using Off-the-Shelf Sensors: Heart Rate Analysis Software
862 from the Taking the Fast Lane Project. *Journal of Open Research Software*, 7, 32.
863 <https://doi.org/10.5334/jors.241>
- 864 Ye, Q., Zou, F., Lau, H., Hu, Y., & Kwok, S. C. (2018). Causal Evidence for Mnemonic
865 Metacognition in Human Precuneus. *The Journal of Neuroscience*, 38(28), 6379–6387.
866 <https://doi.org/10.1523/JNEUROSCI.0660-18.2018>
- 867 Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence
868 and error monitoring. *Philosophical Transactions of the Royal Society B: Biological
869 Sciences*, 367(1594), 1310–1321. <https://doi.org/10.1098/rstb.2011.0416>
- 870 Yonelinas, A. P., & Ritchey, M. (2015). The slow forgetting of emotional episodic memories:
871 An emotional binding account. *Trends in Cognitive Sciences*, 19(5), 259–267.
872 <https://doi.org/10.1016/j.tics.2015.02.009>
- 873