**Differential effects of negative valence and memory type on accuracy, confidence, and metacognitive efficiency**

Juan Castillo1, Patricia Sieweyumptewa1, Elizabeth A. Phelps\*1,2

¹ Department of Psychology, Harvard University

2 Center for Brain Science, Harvard University

**Author Note**

Juan Castillo <https://orcid.org/0000-0001-8964-5187>

Elizabeth A. Phelps  <https://orcid.org/0000-0002-6215-8159>

**Abstract**

Emotion enhances the subjective experience of recollection. Previous research examined associations between memory accuracy and metamemory confidence judgments, but these studies have not quantified the relationship between accuracy and metacognitive confidence judgments. In this study we utilize signal detection theory frameworks to investigate how memory accuracy (measured by discrimination sensitivity) and the alignment between metamemory confidence judgments and memory accuracy (ie. metacognitive efficiency) varies for neutral and negative valence, as well as item and associative detail memory types. Our results indicate that valence and memory type have different effects on accuracy, confidence, and metacognitive efficiency. Negative valence was associated with enhanced accuracy for both items and associated details, but its relationship with response bias varied across memory types, with conservative recognition responses observed for items and liberal responses for associative details. We also observed a double dissociation between metamemory confidence judgements across valence and memory type with negative valence associated with increased confidence for item memory, but decreased confidence for associated details. Examining the association between memory accuracy and metamemory confidence revealed that metacognitive efficiency was greater for negatively valenced items compared to neutral, but this effect did not generalize to details associated with negatively valenced items. These findings advance our understanding of how arousing, negatively valenced information modulates memory and metacognition.

**Keywords:** Metacognition, Memory, Emotion, Confidence

**Introduction**

Highly emotional experiences often result in vivid recollections of the arousing event and its associated details [1]. This is demonstrated by laboratory studies investigating the subjective experience of remembering and the related concept of metamemory confidence, which reflects individuals’ certainty in the accuracy of their recollection [2]. Studies investigating the subjective experience of remembering for affective vs. neutral scenes find that arousing, negatively valenced stimuli are associated with richer self-reported recollective experiences [3] and increased metamemory confidence [4]. Enhanced recollection and confidence correlate with improved memory accuracy for neutrally valenced stimuli [5], but the data for emotional stimuli is mixed. For negatively valenced emotional stimuli, enhanced subjective recollection and confidence is associated with better [3-4, 6] or equivalent [7-8] memory accuracy for items, and better [6], equivalent [4] and worse [4, 9] memory accuracy for associated details. These studies typically compared memory accuracy to participants’ self-reported confidence or recollective experience, but did not precisely quantify the relationship between memory accuracy of metamemory confidence judgements. Moreover, all past studies either did not specifically require participants to report their subjective confidence or failed to account for participants' tendencies to favor one type of response over another (response bias), confidence biases in metacognition, and sequential response biases. In the current study, we leverage signal detection theory frameworks to mitigate the potential biases present in earlier research and quantitatively characterize how the quality of metamemory confidence judgments differs for negatively valenced items and associated contextual details.

The vast majority of previous studies investigating associations between memory accuracy and subjective recollective experience examined the dual processes of recollection and familiarity – mnemonic processes differing in their phenomenological experience due to support from perceptual or semantic information, respectively [10]. Under this framework, recollection, as opposed to familiarity, is considered an enhanced recollective experience. To investigate these mnemonic processes, memory researchers leverage the Remember/Know (R/K) task where participants are presented novel and previously seen stimuli, and provide a ‘remember’ judgment when a stimulus evokes a memory accompanied by episodic details, or a ‘know’ judgment otherwise. This research demonstrates that subjects more accurately recognize items and contextual information for neutrally valenced stimuli they ‘remember’ [5]. However, associations between recollective experience and accuracy for emotional stimuli in the R/K paradigm are more inconsistent. Emotional responses to stimuli presented in the laboratory consistently amplify subjective experiences of recollection, but this is not always accompanied by enhanced memory accuracy for emotional items [8], and can be associated with worse accuracy for associated details [9]. Altogether, these findings suggest that subjective experiences of recollection for emotional events elicited in the lab are partly independent from the accuracy of recollection, and that perceptions of recollective experiences may depend on other factors like vividness or arousal [8, 11].

The experience of recollection compared to familiarity is thought to mirror confidence, but unlike confidence judgments, the R/K paradigm does not explicitly assess whether participants believe their memory is accurate. To our knowledge, only two research groups using the R/K paradigm also investigated how metamemory confidence is associated with memory accuracy. Perfect and colleagues [5] elicited confidence and R/K judgments to evaluate the association between subjective recollective experience and memory accuracy for contextual details. Their investigation revealed that source judgments were more accurate for “remember” and high-confidence responses compared to “know” and low confidence responses. Rimmele and colleagues [4, 9] also probed how confidence and R/K judgments were associated with the accurate recollection of associated details, but for emotional and neutral stimuli. Like the findings of Perfect et al., high confidence and “remember” judgments were similarly associated with detail memory accuracy, however the relationship between remember and high confidence judgments with associated detail accuracy varied depending of type of detail and valence.

Alignment between R/K and confidence judgements is encouraging, but limitations in the R/K framework hinder the generalizability of findings to metamemory confidence judgments. One striking limitation is a lack of consensus among practitioners and laypeople about the meaning of remember and know judgments [12]. This ambiguity casts doubt on the face validity of the R/K task and has prompted researchers to caution against interpreting remember judgments as indicative of high confidence. Roediger & Tekin [13] contend that R/K judgments are not reducible to confidence. They present evidence that participants can report high confidence for both their remember and know judgements, with high confidence remember judgments associated with superior source memory accuracy compared to high confidence know judgments. Another inherent limitation of R/K judgments is their reliance on distinguishing between two qualitative responses to assess differences in subjective recollection – which precludes detailed analyses of incremental differences in recollective experience. Therefore, although research using the R/K paradigm hints at potential associations between metamemory confidence and memory accuracy, results from this paradigm cannot directly inform us about associations between confidence in memory accuracy and memory accuracy itself.

Another issue with previous research investigating metamemory confidence and accuracy concerns biases introduced by sequentially eliciting recognition and confidence judgments – like probing recognition judgments before confidence judgments, or vice versa. A meta-analysis investigating what influences the test-retest reliability of metacognitive performance demonstrates that sequentially eliciting recognition and confidence judgments can undermine reliability. Specifically, test-retest reliability for simultaneous choice and confidence responses is 38.5% higher than when elicited separately (r=0.51; [14]). Simultaneously eliciting choice and confidence judgments is therefore crucial for ensuring the robustness and reproducibility of metamemory research. However, all but one previous study investigating how emotion influences associations between metamemory confidence and accuracy elicited judgments sequentially. This one study demonstrated a double dissociation between subjective recollection and accuracy – recollection for some, but not all, details associated with negatively valenced items is less accurate compared to neutral items for high confidence recognition responses [4]. This study is informative, but still suffers from lingering analytical confounds including the lack of a robust metric for quantifying the relationship between memory accuracy and metamemory confidence judgments.

Addressing the shortcomings present in previous research requires an a priori commitment to frameworks which characterize how biases in recognition, and confidence, are associated with the precision of mnemonic decision making. Signal Detection Theory (SDT) is a computational framework which represents decision-making in the presence of uncertainty as discriminating between meaningful signals (targets) and irrelevant background noise [15]. This framework rationalizes decisions as arising from a generative model with two distinct features: response biases compared to an ideal observer – like always responding yes – and discrimination sensitivity (d’), the approximate separation between target and noise signal distributions [16]. SDT therefore provides a principled framework for modeling the appropriateness of choice behavior while accounting for individual differences in response bias. Absent this framework, memory researchers may misattribute an enhanced sense of recollection for emotional stimuli to improved accuracy instead of a more liberal response bias [17].

In contrast to recognition judgments, which involve decisions based on information that can be environmentally perceived, confidence judgments rely on representational aspects of perception [18]. This difference suggests the need to separately model the sensitivity and bias of metamemory confidence judgments – a conclusion supported by computational research demonstrating that decisions and confidence rely on the same underlying representation, but are maintained partly separately [19]. This reveals a common limitation of previous research examining associations between discrimination sensitivity and confidence – the latter is susceptible to confidence biases [20]. Fortunately, SDT frameworks can be extended to address this limitation and estimate three variables associated with metacognitive confidence: metacognitive *bias,* metacognitive *sensitivity*, and metacognitive *efficiency.* Metacognitive *bias* captures differences in subjective confidence unrelated to task performance – like a tendency to report high confidence independent of accuracy. Metacognitive *sensitivity* (meta-d’) captures the efficacy of distinguishing between incorrect and correct judgments when assigning confidence ratings (e.g. being accurate for high confidence judgments and inaccurate for low confidence). Specifically, meta-d’ is the idealized first-order sensitivity (d’) expected from confidence ratings given a metacognitively ideal observer. Metacognitive *efficiency* (meta-d’/d’), also known as M-Ratio, compares idealized sensitivity (meta-d’) to observed sensitivity (d’) to characterize how well metacognitive confidence ratings align with performance. Metacognitive efficiency is crucial as it enables meaningful comparisons across individuals with different performance, response, and confidence biases. In sum, SDT provides a flexible and robust theory-driven computational modeling framework which utilizes knowledge of the generative processes underlying behavior to effectively summarize choice and metacognitive behavior using a small set of parameters.

Beyond enabling unbiased quantitative relationships between confidence and accuracy to be established the reliability of SDT estimation frameworks has recently been boosted by hierarchical Bayesian methods of parameter estimation for metacognitive sensitivity (H-meta-d’). Hierarchical methods promote generalizable conclusions and require less data to stabilize estimates by incorporating uncertainty in group-level parameter estimates and minimizing the influence of outliers at the individual-level [21]. A hierarchical framework also allows parameter estimation processes to control for potential trait-level confounds via individual-level covariates. Specifically, simultaneous hierarchical estimation of a regression parameter which controls variation in M-ratio values relative to individual-level predictors. This embeds the estimation of trait-cognition relationships into the parameter inference routine to estimate the influence of individual-level trait differences on metacognitive performance [22].

In sum, research suggests the relationship between metamemory confidence and memory accuracy differs between neutral and negatively valenced items, but methodological shortcomings in previous studies limit the robustness and generalizability of these findings. To address these confounds and quantitatively characterize the association between metamemory confidence and memory accuracy, we leverage SDT frameworks in our experiment: a longitudinal 2-day study investigating how metacognitive confidence judgments are associated with the accuracy of memory for items and their associated details across valence. This strategic experimental design enabled us to reliably characterize mnemonic choice behavior, precisely quantify how associations between metamemory confidence and memory accuracy change when exposed to negatively valenced information, evaluate whether the metacognitive efficiency of metamemory confidence judgments is consistent across memory types, and determine the influence of individual-level trait differences on the metacognitive efficiency of item memory.

**Methods**

*Experimental Design*

Participants completed an incidental encoding session for neutral and negatively valenced images followed 24 hours later by a surprise memory test assessing recognition and subjective confidence in recognition for images and an associated contextual detail – a colored border surrounding each image (Yellow, Green, Blue, Red).

During the incidental encoding session participants categorized images and symbols. Image categorization always occurred first. During image categorization participants viewed 60 scenes with colored borders (30 neutral, 30 negative) and responded whether the border color appeared in each image. Image categorization trials consisted of a 4000ms presentation followed by a 1000ms fixation cross. After the image categorization task participants completed a symbol categorization task administered to preclude retrieval practice. During this distractor task participants indicated whether presented stimuli were letters or numbers; each trial consisted of a 4500ms stimulus presentation. We do not discuss results from this task further.

After finishing both tasks, participants completed self-report questionnaires on Qualtrics (<https://harvard.az1.qualtrics.com>). Two questionnaires were selected to characterize trait dimensions which may influence the subjective experience associated with retrieving episodic memories: the 16-item Vividness of Visual Imagery Questionnaire (VVIQ), and the short form of Porges’ Body Perception Questionnaire (BPQ-SF) – a 46 item questionnaire probing body awareness. These characteristics were evaluated because individual differences in body awareness have previously been shown to bias metacognitive efficiency [22], and mental imagery is positively associated with recognition accuracy for colored photographs [23].

Participants returned 24 hours after completing the encoding session for a surprise memory test probing item and associative detail memory. They were notified that the test would include all 60 previously categorized items (old) and 60 novel (new) lures, totaling 120 item recognition judgments. First, participants were presented with an image and asked to report their confidence level (‘not confident’, ‘somewhat confident’, or ‘very confident’) that a given item was ‘old’ or ‘new’. This nested confidence and response scale was designed to minimize biases introduced by sequentially forming confidence judgments after reaching a decision [14, 19]. For self-reported ‘old’ responses, participants were shown the item paired with the associated detail – a colored border. Participants were also notified of lures in this associative detail memory test which used the same nested response scale as the item memory test and required them to report their confidence level that a given old item was paired with old or new colored border. Presented and novel stimuli were counterbalanced across subjects, and border color frequency was equivalent across valence and lure conditions. Responding was self-paced, and a shortened practice version of the task was completed to ensure participants understood the task and response procedure.

*Participants*

Seventy-four participants were recruited from Harvard Universities’ SONA participant recruitment platform (<https://husp.sona-systems.com/>), and completed both parts of this 2-day study. Several participants were excluded from reported analyses due to our preregistered exclusion criteria: not using the complete range of the confidence scale (n=2), failing encoding attention checks (n=5), poor item memory accuracy <70% (n=3), and below chance (50%) memory accuracy for associated details (n=2).

Beyond the preregistered exclusion criteria, three supplemental exclusion criteria were added as unexpected confounds arose. One participant presented with Aphantasia, the inability to perform mental imagery on the VVIQ; we excluded this participant since they markedly differed from the rest of our sample. Additionally, a recent meta-analysis investigating the test-retest reliability of metacognitive performance metrics recommended employing regularization on M-ratio values like a simple bounding method with bounds between 0 and 1.6 [14]. We therefore excluded one participant whose M-ratio fell outside these bounds to improve the reproducibility of our results. Finally, a fundamental prerequisite of estimating Meta-d’ is that individuals must perform better than chance on discrimination sensitivity (d’>0.5), because otherwise individuals cannot be assumed to accurately self-monitor their performance. We ultimately excluded 20 participants in the associative detail memory condition to promote the validity of comparing metacognitive efficiency across memory type. For completeness, we repeated all item memory analyses including this held out sample of participants and report the findings in our supplemental materials (Supplementary Tables S22-S31).

A total of 40 participants (54.1%; 34 Female, 6 Male; Age M = 27.07, SD = 9.85; Race: 19 White (47.5%), 12 Asian (30%), 2 Black (7.5%), 3 Multiracial (15%); Ethnicity: 2 Hispanic (5%)) are included in our final analysis. Given that there was a large difference between our gender distribution we conducted supplemental analyses to confirm gender did not confound our results (Supplementary Tables S32-S49). All participants provided informed consent and ethics approval was obtained in accordance with the guidelines and regulations set forth by the Harvard University Committee on the Use of Human Subjects (IRB19-0789).

*Stimuli*

Presented and novel scenes were drawn from previous studies investigating associations between the subjective experience of remembering and memory accuracy [4, 9]. Scenes were chosen from the International Affective Picture System based on normative ratings for emotional arousal and valence [24]. Presented scenes were divided into emotional (arousal: M = 5.62, SD = .63, valence: M = 2.88, SD = .74) and neutral sets (arousal: M = 3.87, SD = .94, valence: M = 5.58, SD = .6), and did not differ in their visual complexity (Rimmele et al. [4]; see Supplemental Materials). To validate normative arousal and valence ratings an independent experiment (n=22) was carried out where participants rated the selected scenes on valence and arousal. Ratings from our subjects confirmed the expected trends with a 100% probability of valence being higher for neutrally valenced items (*M* = 3, *HDI89%* = [2.67, 3.36]), and 100% probability of arousal being greater for negatively valenced items (*M* = 2.26, *HDI89%* = [1.81, 2.7]).

**Statistical Analysis**

*Bayesian Linear Models*

Differences in signal detection theory derived parameters – discrimination sensitivity (d’) and response bias (c) – were evaluated using Bayesian mixed-effects linear models implemented in R and interfaced with Stan via the *brm* function in the ‘*brms*’ package [25]. Parameters estimated under signal detection theory are standardized values derived from z-scores. This standardization leads us to expect that variability around estimated sample means will follow a normal distribution. However, methodological limitations in previous research limit reliable knowledge about anticipated sample means. Consequently, we employ wide gaussian priors covering the likely range of potential values across all our linear models (See Supplementary Methods). This approach reflects our current uncertainty regarding prior knowledge and integrates it into the inferential process. It also ensures that data-driven evidence (likelihood) plays a predominant role in our parameter estimation process. See our supplementary materials for prior specification information (Supplementary Table S1).

Linear models estimating within-subjects differences in valence and memory type conditions included random effects at the participant level.Leveraging Bayesian models allowed us to take advantage of several Bayesian methods to confirm the robustness of our findings including model comparisons using Bayes factors, and the calculation of posterior Highest Density Intervals (HDI) and posterior probability of direction (pd). Whereas Bayes factors assist in arbitrating between models based on the evidence provided by the data, HDI and pd help interpret the significance of estimated parameters. In essence, an HDI estimates the uncertainty of estimated parameter values and is similar to the frequentist concept of confidence intervals. For example. when the estimated value of a parameter does not include 0 in an 89% HDI, we can say with 89% certainty that the true value of the parameter is not 0. Such effects are considered meaningful, analogous to frequentist conceptualizations of statistical significance. In contrast, the posterior probability of direction (pd) indicates how certain we can be about the estimated directionality of an effect, such as being negatively or positively associated with the dependent variable. Simulation-based evidence indicates that pd possesses the attractive utility of having a direct 1:1 correspondence with frequentist *p*-values [26]. Thus, pd can be interpreted as an index of effect existence, rather than statistical significance, which is approximately equivalent to 1 – *p*-value for a one-sided test*.* Unless otherwise noted, Bayesian statistics are presented with median, HDI calculated at 89% threshold, and posterior probability of direction. We avoid using a 95% threshold for parameter interpretation as this interval choice is traditionally based on frequentist a-priori justification processes balancing alpha error and statistical power – considerations not applicable in a Bayesian framework. Additionally, 89% intervals have been empirically demonstrated to be more stable and reliable than 95% intervals, offering more realistic representations of uncertainty in parameter estimates with limited samples [27]. See our supplemental materials for detailed MCMC sampling summaries (Supplementary Tables S2-S21).

*Signal Detection Theory*

We extracted individual level SDT parameters to evaluate how recognition behavior differs across memory type, valence, and is associated with individual traits including visual imagery and body awareness. In maximum likelihood estimation, edge correction is traditionally applied to correct for zero counts in false alarms, correct rejections, hits, and misses [28]. However, this practice strongly biases individual-level parameter estimates in situations with limited trial counts. Hierarchical estimation applies a shrinkage effect on individual-level estimates by sharing information across individual and group levels, but outputs more accurate and generalizable estimates. Given the limited number of trials feasible for a memory experiment, we extract individual-level estimates from our hierarchical estimation process.

*H-Meta-d’*

We apply a hierarchical Bayesian framework to obtain reliable estimates of metacognitive efficiency at the group level (H-Meta-d’). Data were fit according to the specifications necessary for analysis in MATLAB with the H-Meta-d’ toolbox (<https://github.com/metacoglab/HMeta-d>), and instructions detailing these components are described elsewhere [21]. The H-Meta-d’ toolbox uses Markov Chain Monte Carlo (MCMC) sampling to estimate posterior distributions over model parameters using JAGS. See supplementary materials for more information regarding MCMC sampling specifications. Hierarchical sampling enables accurate estimation of group-level parameters by allowing group-level estimates to constrain individual-level fits – minimizing the impact of individual estimates with high uncertainty at the group-level (ie. outliers). Group-level metacognitive performance was compared across neutral and negatively valenced items, and for items and their associated details. The significance of parameter values was determined by evaluating whether 89% HDIs on the posterior distributions included zero.

*RH-Meta-d’*

We further extend our hierarchical framework for estimating metacognitive efficiency to estimate regression parameters – the specifics of which are described in detail elsewhere [22]. This extension enables inferences about whether individual traits influence metacognitive efficiency at the group-level.

**Results**

We first report memory and metacognitive performance for item recognition collapsed across valence. Next, we examine the relationship between overall item memory accuracy and visual imagery, as well as body awareness. We then describe how memory and metacognitive performance vary across valence for item recognition. Subsequently, we outline the differences in memory and metacognitive performance between items and their associated details. Finally, we report how memory and metacognitive performance differ across valence for associated details.

*Recognition judgments for items were liberal (c), but demonstrated high discrimination sensitivity (d’)*

Our sample performed well on the old/new recognition task – effectively discriminating between previously seen and novel items (*M* = 2.23, *SD* = 0.32). This performance is particularly noteworthy given liberal response biases demonstrated by an increased tendency to claim recognition compared to an unbiased observer (*M* = -0.36, *SD* = 0.11). Alignment between effective discrimination and liberal responding suggests participants were metacognitively aware of their performance. However, we must evaluate confidence judgments and metacognitive efficiency to confirm this association.

*Metamemory confidence judgments for items were metacognitively inefficient*

Average confidence in our sample was 2.45 (*SD* = 0.26). Compared to the median of our confidence scale (2), this indicates participants in our sample were overconfident. To evaluate how well metamemory confidence ratings align with accuracy, we hierarchically estimated metacognitive efficiency at the group level while taking into consideration uncertainty at the individual level. Metacognitively ideal observers align confidence ratings with performance and possess an M-ratio of 1. In our sample, metamemory confidence judgments were metacognitively inefficient (*Median* = 0.55, *SD* = 0.097). Considering participant’s high discrimination sensitivity, this suggests metacognitive inefficiency stemmed from overconfidence in performance.

*Visual imagery and body awareness are not associated with recognition or metacognitive judgments for item memory*

Our participants possessed a wide range of individual differences with respect to body awareness (*M* = 65.9, *SD* = 21.76, Range=29-110), and visual imagery capabilities (*M* = 57.9, *SD* = 10.26, Range=41-80). We investigated potential associations between SDT parameters (c/d’) and these trait-level features for item memory to determine if individuals with greater awareness of their physical and mental processes leverage these capabilities in service of recognition. Specifically, we leveraged Bayesian linear models with visual imagery and body awareness as independent variables, and SDT parameters d’ and c as our dependent variables. There was a lack of evidence to support meaningful associations across all four analyses (Supplemental Tables S8-S11). We additionally investigated whether visual imagery or body awareness is associated with metacognitive efficiency. Specifically, we follow a procedure from previous research [22] to conduct a hierarchical regression estimate of metacognitive efficiency where participant’s standardized visual imagery and body awareness subscale scores serve as our covariates of interest. This analysis concluded there is insufficient evidence to support an association between visual imagery and metacognitive efficiency with 89% confidence (*Median* = -0.13, HDI89%[-0.3 0.032]), and insufficient evidence to support an association between body awareness and metacognitive efficiency with 89% confidence (*Median* = 0.11, HDI89%[-0.051, 0.27]). Therefore, the evidence indicates that neither visual imagery nor body awareness is positively associated withrecognition or metacognitive judgmentsfor items.

*Recognition judgments for negatively valenced items are more conservative, and sensitive, than for neutral items*

Recognition judgments were more sensitive (d’: *M*=2.44, *SD*=0.62), and conservative (c: *M*=-0.2, *SD*=0.27) for negatively valenced items compared to neutral (d’: *M*=2.02*, SD*=0.46*;* c: *M*=-0.43, *SD*=0.26). We can be at least 89% certain that discrimination sensitivity meaningfully differs across valence (*M* = 0.37, *HDI89%* = [0.23, 0.51]) with a 99.99% probability of being higher for negatively valenced items, and response bias meaningfully differs across valence (*M* = -0.18 *HDI89%* = [-0.26, -0.09]) with a 99.94% probability of being more conservative for negatively valenced items. Differences in behavior across valence conditions suggest that individuals increased discrimination sensitivity and conservative response behavior was due to the arousing, negatively valenced nature of emotional stimuli.

*Metacognitive efficiency was 1.16x greater for negatively valenced items*

Average confidence was high in our sample, but qualitative differences across valence conditions were still apparent; metamemory confidence judgments for negatively valenced items were higher (*M*=2.49, *SD*=0.27) than neutral (*M*=2.41, *SD*=0.27). We can be at least 89% certain that confidence ratings meaningfully differ across valence (*M* = 0.08, *HDI89%* = [0.05, 0.12]) with a 99.94% probability of being higher for negatively valenced items. This conclusion is additionally supported by 75% (30/40) of participants being more confident when recognizing negatively valenced items.

The metacognitive efficiency of metamemory confidence judgments was higher for negatively valenced items (*M*=0.74, *SD*=0.079) compared to neutral (*M*=0.64, *SD*=0.16). Comparing metacognitive efficiency across valence conditions in item memory provides evidence that we can be at most 88% certain there is an effect of valence on metacognitive efficiency (*Median* = 0.13, HDI89%[0.001, 0.26]). The evidence therefore indicates that negative valence is associated with improvements in metacognitive efficiency for item memory, but further investigations are necessary to replicate this finding and evaluate whether this association generalizes to other features of negatively valenced episodic memories like associated details.

*Recognition judgments for associated details are more conservative and less sensitive than for items*

Recognition judgments for items were highly sensitive (d’: *M*=2.22, *SD*=0.32), and liberal (c: *M*=-0.36, *SD*=0.14) compared to associated details (d’: *M*=1.3*, SD*=0.45*;* c: *M*=0.35, *SD*=0.064). We can be at least 89% certain that discrimination sensitivity meaningfully differs across memory type (*M* = -0.89, *HDI89%* = [-1.03, -0.75]) with a 100% probability of being lower for associated details, and response bias meaningfully differs across memory type (*M* = 0.71, *HDI89%* = [0.67, 0.74]) with a 100% probability of being more conservative for associated details. Alignment between diminished discrimination and conservative responding for associative details suggests participants were metacognitively aware of their performance. However, we must evaluate confidence judgments and metacognitive efficiency to confirm this association.

*Metacognitive efficiency was 3.06x greater for items compared to associated details*

Metamemory confidence judgments were higher for items (*M*=2.45, *SD*=0.26) compared to associated details (*M*=1.91, *SD*=0.36). We can be at least 89% certain that confidence ratings meaningfully differ across memory type (*M* = 0.54, *HDI89%* = [0.45, 0.63]) with a 100% probability of being higher for items. This conclusion is additionally supported by trends demonstrating that 97.5% (39/40) of participants were more confident in their recognition of items than associated details.

The metacognitive efficiency of metamemory confidence judgments for items (*Median*=0.52, *SD*=0.12) was qualitatively greater than the metacognitive efficiency of associated details (*Median*=0.17, *SD*=0.24). Comparing metacognitive efficiency across memory type conditions provides evidence that we can be at least 89% certain there are differences in metacognitive efficiency across memory type (*Median* = 0.37, HDI89%[0.24, 0.49]). This conclusion is additionally supported by 92.5% (37/40) of participants being more metacognitively efficient for item metamemory judgments. This provides compelling evidence that the metacognitive efficiency of item metamemory confidence judgments is more than three times greater than associated details.

*The effect of negative valence on metamemory confidence judgments differs by memory type*

Given the observed differences in confidence ratings across valence and memory type, we examined whether an interaction effect existed between these factors. Towards this end, we strategically arbitrated between potential model specifications for this multivariate analysis by using Bayes Factor (BF) model comparisons. This analysis provided moderate support (*BF10 = 3*) for an interaction effect between valence and memory type as the best fitting model. Results from this interaction model indicate we can be at least 89% certain that the effect of negative valence on confidence is meaningful (*M* = 0.08, *HDI89%* = [0.01, 0.16]) with a 95.44% probability of being positive – indicating an expected increase for negatively valenced item metamemory confidence ratings compared to neutral. Additionally, we can be at least 89% certain that the effect of memory type on confidence is meaningful (*M* = -0.44, *HDI89%* = [-0.52, -0.36]) with a 100% probability of being negative for associative memory – indicating individuals were less confident in remembering the details associated with the items they previously saw. Critically, we can be at least 89% certain that the interaction term effect between memory type and valence is meaningful (*M* = -0.19, *HDI89%* = [-0.31, -0.08]) with a 99.72% probability of being negative. This means that compared to item memory, the effect of negative emotional valence on confidence ratings is reversed for associated details. Specifically, the effect of negative valence on metamemory confidence judgments differs by memory type – increasing metamemory confidence judgments for item memory, but decreasing them for associated details (Figure 1).

A graph with blue and red lines

Description automatically generated

Figure 1. Interaction effect between valence and memory on metamemory confidence judgments

*Recognition judgments for details associated with negatively valenced items are less conservative, and more sensitive, than for neutral*

The discrimination sensitivity (d’) of recognition judgments was greater for details associated with negatively valenced items (*M*=1.35, *SD*=0.74) compared to neutral (*M*=1.17, *SD*=0.77), and recognition was less conservative for details associated with negatively valenced items (*M*=0.24, *SD*=0.31) compared to neutral (*M*=0.42, *SD*=0.3). We can be at most 88% certain that discrimination sensitivity for associated details differs across valence (*M* = -0.19, *HDI89%* = [-0.37, 0.00]) with a 93.75% probability of being higher for details associated with negatively valenced items, and at least 89% certain that response bias meaningfully differs across valence (*M* = 0.18, *HDI89%* = [0.08, 0.28]) with a 99.97% probability of being less conservative for details associated with negatively valenced items. Our results therefore provide evidence that negative valence improves discrimination sensitivity for items and their associated details, but differs regarding its effect on response bias – resulting in conservative recognition judgments for items and liberal recognition for associative details.

*Metacognitive efficiency does not differ across valence for associated details*

The metacognitive efficiency of metamemory confidence judgments for details associated with negatively valenced items (*M*=0.61, *SD*=0.056) was qualitatively lower compared to neutral items (*M*=0.83, *SD*=0.15). Comparing metacognitive efficiency across valence for associative details provided insufficient evidence to assert that valence effects the metacognitive efficiency of associated details with 89% confidence (*Median* = -0.17, HDI89%[-0.45, 0.14]). Therefore, negative valence improves metacognitive efficiency for item memory, but this effect does not extend to associated details.

**Discussion**

Our study demonstrates that item recognition is both highly sensitive and liberal, especially when compared to the recognition of details associated with items. Discrimination sensitivity for items was improved for emotional images – in alignment with research demonstrating improved recognition for arousing, negatively valenced stimuli [3, 4, 6]. However, our results diverge from earlier studies linking negative information in item memory with liberal recognition judgments [17]. Instead, we observed more conservative response biases for the recognition of negatively valenced items. One potential explanation for this divergence is that different types of stimuli were leveraged as items. In contrast to words [17], emotionally evocative images may have evoked different response strategies.

Our results also demonstrate that negative valence is associated with improved discrimination sensitivity, but not metacognitive efficiency, for details associated with negatively valenced items. This pattern of results resembles findings from Rimmele et al [4] which demonstrate that memory accuracy for associated item details (location and temporality) was improved for negatively valenced items, whereas differences across low and high confidence for details associated with negatively valenced items failed to reach significance. The improved accuracy of memory for details associated with negatively valenced items is consistent with other studies as well [6, 29], but not all research investigating memory accuracy for contextual details [4, 9, 30]. Rimmele et al. [9] also evaluated memory for the same item and associative detail as our work (ie. images paired with colored borders), but did not observe improvements in memory for details associated with negatively valenced items; however, comparisons to this study may not be valid as memory accuracy was characterized by evaluating high confidence and remember hit responses – which do not control for confidence or response biases. Notably, Rimmele et al. [4, 9] was the only research group we identified which elicited recognition and confidence responses simultaneously. This raises the concern that contrasting results from other studies may have arisen from response biases, sequential responding or differences in the type of associative detail investigated. By isolating our investigation to a single associated detail and controlling for biases with strategic experimental design, our study reveals that negative valence improves discrimination sensitivity for associated details. However, one possibility is that this result may have arisen due to the selection process we enforced to promote reliable comparisons of metacognitive efficiency across memory type (d’ > 0.5).

Our investigation is the first we are aware of which quantitatively characterizes the metacognitive efficiency of metamemory confidence judgments – the accuracy of metacognitive judgments regarding recognition performance. On average metacognitive efficiency was 0.51, generally inefficient compared to an ideal Bayesian observer. Notably, metamemory confidence judgments were 3.06x greater for items compared to their associated details. This is particularly striking because we observed a meaningful average decrease of 0.51 in confidence judgments for details associated with items. Decreased metamemory confidence and increased conservative recognition biases for details compared to items suggests participants were metacognitively aware of the inherent uncertainty involved in recalling specific details associated with items; however, their poor metacognitive efficiency underscores a failure to effectively monitor this uncertainty.

Arousing, negatively valenced items were associated with 1.16x greater metacognitive efficiency compared to neutral stimuli, but this effect did not extend to details associated with items. The selective effect of arousing, negatively valenced information on the metacognitive efficiency of items, but not their associated details, signals the need for more extensive investigations of how metacognitive efficiency changes across valence for different memory types. Our investigation focused on memory for associated details unrelated to the core image content (ie. peripheral border color). However, previous research indicates that recognition and remember judgments are more accurate for intrinsic details central to an item's identity, such as its color properties [31, 32]. By leveraging the signal detection frameworks utilized in this article, researchers can reliably characterize fine-grained distinctions in the metacognitive efficiency of metamemory confidence judgments and determine whether differences exist for various types of mnemonic details.

Extensive research links the subjective vividness of emotional memories to the perceived accuracy of these memories [8, 30]. However, previous efforts failed to characterize this relationship with quantitative precision. By utilizing signal detection frameworks to estimate discrimination sensitivity, response bias, and metacognitive efficiency, our analysis provides the novel insight that arousing, negatively valenced information is positively associated with increased discrimination sensitivity and metacognitive efficiency for metamemory confidence judgments. This study is the first to precisely delineate the latter association – notably demonstrating its specificity for the memory of items over their associated details. Heightened discrimination sensitivity for negatively valenced items and their associated details supports propositions by Phelps & Sharot [8]. Specifically, they proposed that an enhanced subjective sense of recollection for emotionally charged memories likely stems from recalling a single item or specific details associated with an emotional event. However, we refine this perspective by demonstrating that enhanced recollective experiences for emotional events are only accurate for items, not their associated details. This insight highlights a potential evolutionary advantage conferred by metacognition for decision-making involving arousing, negatively valenced items, and motivates future investigations to characterize this association.

**Acknowledgments**

We would like to thank Dr. Stephen Fleming for perspectives on experimental design and Dr. Daniel L. Schacter for insights on an earlier version of this manuscript.

**Author Contributions**

JC and EAP were involved in conceptualization, methodology, writing, supervision, and project administration. EAP was involved in funding acquisition. JC was involved in visualization. JC and PS were involved in investigation, data curation, formal analysis, and project administration.

**Data Availability**

We report how we determined participant exclusion criteria, manipulations and measures, and the software used for analyses. Additionally, to promote the quality and transparency of our analyses, we follow several Bayesian Analysis Reporting Guidelines - a standardized list of criteria for producing and reporting reproducible Bayesian analyses for scientific publication [33]. Analysis code, experimental data, and MCMC sampling data is available at <https://github.com/DaPsyientist/Metad>. This study’s design and analysis were pre-registered at <https://aspredicted.org/ZVH_6FM>.

**Competing Interests**  
The authors declare no competing interests.

**References**

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

[2] Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. *American Psychologist,* ***34***(10)**,** 906-911. <https://doi.org/10.1037/0003-066x.34.10.906>

[3] Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General,* ***129***(2)**,** 242-261. <https://doi.org/10.1037/0096-3445.129.2.242>

[4] Rimmele, U., Davachi, L., & Phelps, E. A. (2012). Memory for time and place contributes to enhanced confidence in memories for emotional events. *Emotion,* ***12***(4)**,** 834-846. <https://doi.org/10.1037/a0028003>

[5] Perfect, T. J., Mayes, A. R., Downes, J. J., & Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory. *Quarterly Journal of Experimental Psychology,* ***49***(3)**,** 797-813. <https://doi.org/10.1080/713755644>

[6] Kensinger, E. A., & Corkin, S. (2003). Memory enhancement for emotional words: are emotional words more vividly remembered than neutral words? *Memory & Cognition,* ***31***(8)**,** 1169-1180. <https://doi.org/10.3758/bf03195800>

[7] Sharot, T., Delgado, M. R., & Phelps, E. A. (2004). How emotion enhances the feeling of remembering. *Nature Neuroscience,* ***7***(12)**,** 1376-1380. <https://doi.org/10.1038/nn1353>

[8] Phelps, E. A., & Sharot, T. (2008). How (and why) emotion enhances the subjective sense of recollection. *Current Directions in Psychological Science,* ***17***(2)**,** 147-152. <https://doi.org/10.1111/j.1467-8721.2008.00565.x>

[9] Rimmele, U., Davachi, L., Petrov, R., Dougal, S., & Phelps, E. A. (2011). Emotion enhances the subjective feeling of remembering, despite lower accuracy for contextual details. *Emotion,* ***11***(3)**,** 553-562. <https://doi.org/10.1037/a0024246>

[10] Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language,* ***46***(3)**,** 441-517. <https://doi.org/10.1006/jmla.2002.2864>

[11] Rimmele, U., Lackovic, S. F., Tobe, R. H., Leventhal, B. L., & Phelps, E. A. (2016). Beta-adrenergic blockade at memory encoding, but not retrieval, decreases the subjective sense of recollection. *Journal of Cognitive Neuroscience,* ***28***(6)**,** 895-907. <https://doi.org/10.1162/jocn_a_00941>

[12] Umanath, S., & Coane, J. H. (2020). Face validity of remembering and knowing: empirical consensus and disagreement between participants and researchers. *Perspectives on Psychological Science,* ***15***(6)**,** 1400-1422. <https://doi.org/10.1177/1745691620917672>

[13] Roediger, H. L., & Tekin, E. (2020). Recognition memory: Tulving’s contributions and some new findings. *Neuropsychologia,* ***139,***107350. <https://doi.org/10.1016/j.neuropsychologia.2020.107350>

[14] Guggenmos, M. (2021). Measuring metacognitive performance: type 1 performance dependence and test-retest reliability. *Neuroscience of Consciousness,* ***2021***(1)**,** niab040. <https://doi.org/10.1093/nc/niab040>

[15] Anderson, N. D. (2015). Teaching signal detection theory with pseudoscience. *Frontiers in Psychology,* ***6***. <https://doi.org/10.3389/fpsyg.2015.00762>

[16] Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods Instruments & Computers,* ***31***(1)**,** 137-149. <https://doi.org/10.3758/bf03207704>

[17] Dougal, S., & Rotello, C. M. (2007). “Remembering” emotional words is based on response bias, not recollection. *Psychonomic Bulletin & Review,* ***14***(3)**,** 423-429. <https://doi.org/10.3758/bf03194083>

[18] Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children’s knowledge about visual perception: further evidence for the level 1–level 2 distinction. *Developmental Psychology,* ***17***(1)**,** 99-103. <https://doi.org/10.1037/0012-1649.17.1.99>

[19] Fleming, S. M., & Daw, N. D. (2017). Self-evaluation of decision-making: A general Bayesian framework for metacognitive computation. *Psychological Review,* ***124***(1)**,** 91-114. <https://doi.org/10.1037/rev0000045>

[20] Fleming, S. M., & Lau, H. (2014). How to measure metacognition. *Frontiers in Human Neuroscience,* ***8,*** 443. <https://doi.org/10.3389/fnhum.2014.00443>

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

[22] Harrison, O. K., et al. (2021). The filter detection task for measurement of breathing-related interoception and metacognition. *Biological psychology,* ***165*,** 108185. <https://doi.org/10.1016/j.biopsycho.2021.108185>

[23] Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology,* ***64***(1)**,** 17-24. <https://doi.org/10.1111/j.2044-8295.1973.tb01322.x>

[24] 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>

[25] Bürkner, P.-C. (2017). Brms: an R package for Bayesian multilevel models using stan. *Journal of Statistical Software,* ***80***(1)**,** 1-28. <https://doi.org/10.18637/jss.v080.i01>

[26] Makowski, D., Ben-Shachar, M. S., Chen, S. H. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology,* ***10,*** 2767. <https://doi.org/10.3389/fpsyg.2019.02767>

[27] Kruschke, J. K. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan.* Academic Press.

[28] Hautus, M. J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of d’. *Behavior Research Methods Instruments & Computers,* ***27***(1)**,** 46-51. <https://doi.org/10.3758/bf03203619>

[29] Doerksen, S., & Shimamura, A. P. (2001). Source memory enhancement for emotional words. *Emotion,* ***1***(1)**,** 5-11. <https://doi.org/10.1037/1528-3542.1.1.5>

[30] Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science,* ***14***(5)**,** 455-461. <https://doi.org/10.1111/1467-9280.02453>

[31] Mather M. (2007). Emotional arousal and memory binding: an object-based framework. *Perspectives on psychological science : a journal of the Association for Psychological Science,* ***2***(1)**,** 33–52. <https://doi.org/10.1111/j.1745-6916.2007.00028.x>

[32] Mihaylova, M., Vuilleumier, P., & Rimmele, U. (2019). Better memory for intrinsic versus extrinsic details underlies the enhanced recollective experience of negative events. *Learning & Memory,* ***26***(11)**,** 455-459. <https://doi.org/10.1101/lm.049734.119>

[33] Kruschke, J. K. (2021). Bayesian analysis reporting guidelines. *Nature Human Behaviour,* ***5***(10)**,** 1282–1291. [https://doi.org/10.1038/s41562-021-01177-7](https://psycnet.apa.org/doi/10.1038/s41562-021-01177-7)