Integration of Emotional and Attractiveness Cues Enhances Learning Judgment Accuracy: Evidence from LPP and NSW

Authors: Jiang Yingjie, Guo Yanlin, Zhang Xiaojing, Ren Jimei, Jiang Yingjie

Date: 2024-05-09T00:00:00+00:00

Abstract

Emotion and attractiveness are two important cues that influence learning judgments of face memory. Previous studies have only examined the effects of the two cues on learning judgments separately, neglecting the role of cue integration. However, in real-world situations, emotion and attractiveness typically jointly influence individuals' processing of a single face. The present study investigated the promoting effect of cue integration of emotion and attractiveness on learning judgment accuracy and its electrophysiological mechanisms through two experiments. The results revealed that: (1) More than half of the participants integrated the two cues of emotion and attractiveness for learning judgment ratings. (2) Cue integration of emotion and attractiveness can promote learning judgment accuracy. (3) Compared with the non-cue-integration group, the cue-integration group exhibited larger amplitudes of the late positive potential (LPP) in the parietal region during the encoding stage and the late negative wave (NSW) in the frontal region during the learning judgment stage, both of which were significantly correlated with learning judgment accuracy. The findings indicate that cue integration of emotion and attractiveness can promote learning judgment accuracy, and its underlying mechanism is reflected in better allocation of cognitive resources for cognitive evaluation during the late encoding stage and better information maintenance during the learning judgment stage, thereby enabling more accurate monitoring of one' s own cognitive processes of face memory.

Full Text

Preamble

Self-Check Report for *Acta Psychologica Sinica* Submission

Please complete the following items and paste them on the first page of your manuscript.

1. Please list up to three innovative contributions of this study in the form of "Research Highlights," with a total word count not exceeding 200 words.

Acta Psychologica Sinica aims to publish cutting-edge psychological research that is "both scientifically excellent and of particularly broad interest and significance." If your study only makes minor incremental contributions, does not attempt to open new areas of inquiry or propose unique and innovative perspectives, or particularly if it merely investigates algorithms or techniques without addressing clear psychological questions, such research has a low probability of being accepted by this journal and we recommend submitting elsewhere.

Answer: This study examined face memory and metamemory monitoring as core tasks, investigating how the integration of emotion and attractiveness cues affects judgment-of-learning (JOL) accuracy. Results revealed individual differences in cue integration and demonstrated that cue integration facilitates JOL accuracy. By collecting ERPs data during encoding, JOL, and retrieval phases, we specifically examined the mechanism through which cue integration promotes JOL accuracy, which manifests as better allocation of cognitive resources for cognitive evaluation during the late encoding stage and better information maintenance during the JOL stage, thereby enabling more accurate monitoring of one' s own face memory cognitive processes.

2. Have you published or submitted any articles using the same data as this study? If yes, please attach them for review. (We do not encourage authors to publish multiple articles using the same data with identical variables, nor do we support splitting a series of related studies into multiple publications.)

Answer: No.

3. For non-experimental, non-intervention studies in management, clinical, personality, and social psychology that rely solely on self-report (questionnaire methods), it is necessary to check for common method bias. What methods did you use to control for or demonstrate that such bias would not affect the validity of your conclusions? What measures were taken? (Relevant literature on common method bias can be found at: http://journal.psych.ac.cn/xlkxjz/CN/abstract/abstract894.shtml) Studies based on cross-sectional data with only self-reports and convenient sampling are easy to conduct but typically have limited innovative value and low acceptance probability.

Answer: This study is an experimental study and does not have the above issues.

4. Are effect sizes reported and analyzed (e.g., Cohen' s d for t-tests, ^{2} or ^{2}p for ANOVA, standardized regression coefficients)? (Many studies mechanically report effect sizes without necessary analysis or explanation, such as whether the effect size is small/medium/large or its theoretical/applied significance.) (Searching "effect size calculator" on Google yields many convenient apps. For explanations of effect sizes in Chinese, see: http://journal.psych.ac.cn/xlkxjz/CN/abstract/abstract1150.shtml; in English, see: http://www.uccs.edu/lbecker/effect-size.html) Are 95% CIs reported for statistical analyses? (e.g., 95% CI for differences, 95% CI for correlation/regression coefficients) For calculations and plotting of confidence intervals, see https://thenewstatistics.com/itns/esci/

5. Please state the planned sample size and actual sample size. If they differ, please explain why. Previous psychological research has universally suffered from low statistical power due to insufficient sample sizes. We recommend explaining in the Methods section the basis for your sample size calculation and determination. Sample size should be determined based on a justified effect size and desired power, with the calculation software or program reported. For rationale and practices regarding sample size planning, see https://osf.io/5awp4/

Answer: This study simultaneously examined memory and metamemory monitoring. When determining the number of subjects, we referenced previous studies investigating the effects of cues on memory and metamemory monitoring: Legrand et al. (2021) recruited 35 subjects and retained 30 valid subjects. Ultimately, in this study, Experiments 1 and 2 selected 41 and 45 subjects respectively, retaining 40 and 42 valid datasets.

6. In hypothesis testing, if using Null Hypothesis Significance Testing (NHST), precise p-values must be reported rather than p-value ranges (report ranges only for $p < 0.001$, otherwise report precise p-values). Does your paper meet this requirement? If using Bayes factors, have you reported its sensitivity to prior distribution assumptions?

7. To ensure completeness of data reporting, if some data were excluded from statistical analysis, is this reported in the text? What were the reasons? How would statistical results change if this portion of data were included? How were missing data handled in statistical analysis? When using scales, were any individual items deleted? Why? How would statistical results change if these items were included? Were there any measured items or variables not reported? Why? Please indicate where in the paper this is addressed.

Answer: Yes. Reasons for excluding behavioral experiment subjects: Some subjects withdrew midway due to personal reasons. Reasons for excluding ERPs experiment subjects: When excluding trials with absolute EEG voltage exceeding $±$100 µV, subjects with more than 35% of trials excluded were deemed ineligible. Both have been reported in the text. Statistics···

8. For experimental materials, scales, or questionnaires that have not undergone peer review and examination, are they attached at the end of the document for review? If not, please explain why. If this article is published, are you willing to share these materials with other researchers?

Answer: No such materials were used.

9. This journal requires authors to provide raw data. Please choose one of the following options:

a) Send data to editorial office email after submission

b) Data can be obtained from the following link

c) Raw data and programs have been shared on the Psychological Science Data Bank (https://psych.scidb.cn/)

d) If unable to provide, please explain reasons or provide relevant proof.

Answer: Cannot provide temporarily, as the data in this article are from an unpublished thesis.

10. Is your study a clinical intervention or laboratory experiment? Yes(√) No( )

If yes, please provide pre-registration number. If no, please explain why.
Note: Clinical interventions or laboratory experiments should be pre-registered before data collection. Other experimental studies are also encouraged to pre-register. Pre-registration requires stating all research hypotheses and their support, as well as detailed experimental/intervention procedures. This journal' s pre-registration website is https://os.psych.ac.cn/preregister (see "Download Center" on the journal website for instructions) or https://osf.io/ or https://aspredicted.org/. If your study is pre-registered, it will significantly increase acceptance probability. The importance of pre-registration can be found at https://osf.io/5awp4/

Answer: No, this experiment was conducted 3 years ago, before *Acta Psychologica Sinica* had a pre-registration system.

11. If your study used human or animal subjects, was it approved by your institution' s ethics committee? If yes, please send a scanned copy to the editorial office email. If no, please explain why.

12. Have you written a 400-500 word English abstract following the "English Abstract Writing Guidelines" published on the editorial office

website? Has the English title and abstract been reviewed by English-proficient professionals or edited by a professional SCI/SSCI paper editing company?

Answer: Yes, it has been written and reviewed.

13. If the first author is a student, please have the advisor send a separate email to the editorial office (xuebao@psych.ac.cn) stating that they have read this paper and carefully reviewed it. Have you reminded your advisor to send this email? (The editorial office will only consider entering the manuscript processing workflow after receiving the advisor' s email)

14. Please go to the "Download Center" on the right side of the editorial office website homepage to download and complete the "Manuscript Non-Confidentiality Certificate," stamp it with the confidentiality office seal of the corresponding author' s institution, and send a scanned copy to the editorial office email (xuebao@psych.ac.cn). If there is no confidentiality office seal, please stamp it with the official seal of the corresponding author' s institution. Has the email been sent?

Answer: No, will send to editorial office email after submission.

Cue Integration of Emotion and Attractiveness Facilitates Judgment-of-Learning Accuracy: Evidence from LPP and NSW

Abstract

Emotion and attractiveness are two important cues that influence judgments of learning (JOLs) for face memory. Previous studies have only examined the effects of each cue on JOLs in isolation, neglecting the role of cue integration. However, in real-world situations, emotion and attractiveness typically jointly influence the processing of a single face. The present study investigated how the integration of emotion and attractiveness cues facilitates JOL accuracy and its electrophysiological mechanisms through two experiments. The findings revealed: (1) Over half of the participants integrated emotion and attractiveness cues when making JOL ratings. (2) Integrating emotion and attractiveness cues promoted JOL accuracy. (3) Compared to the non-integration group, the cue integration group exhibited larger amplitude of the late positive potential (LPP) in the parietal region during the encoding stage and larger amplitude of the negative slow wave (NSW) in the frontal region during the JOL stage, both of which were significantly correlated with JOL accuracy. These results demonstrate that cue integration of emotion and attractiveness facilitates JOL accuracy, with the underlying mechanism manifesting as better allocation of

cognitive resources for cognitive evaluation during the late encoding stage and better information maintenance during the JOL stage, thereby enabling more accurate monitoring of one' s own face memory cognitive processes.

Keywords: judgments of learning, cue integration, metamemory

Judgments of learning (JOLs) refer to learners' subjective predictions about whether they can successfully retrieve previously studied material in future tests and represent one of the most commonly used indicators of metamemory monitoring [?]. The relationship between JOLs and actual memory performance is known as JOL accuracy [?]. Accurate JOLs can promote learning and enable individuals to effectively regulate and control their own cognitive processes [?].

Research has shown that emotion and attractiveness are important cues that influence JOLs for face memory [?, ?]. Emotional faces receive higher JOL ratings than neutral faces. Michaela et al. (2020) asked participants to memorize negative and neutral emotional pictures and then make JOLs. Results showed that participants gave higher JOL ratings to negative pictures compared to neutral pictures. Hourihan and Bursey (2017) used positive and neutral emotional pictures as learning materials and similarly found that individuals gave higher JOL ratings to positive pictures than to neutral pictures. In addition to facial emotional information, attractiveness is also an important cue for face JOLs [?]. Researchers manipulated the attractiveness level of faces and found that individuals showed higher JOL accuracy for high and medium attractiveness faces compared to low attractiveness faces [?].

However, the aforementioned studies only examined the effects of facial emotion or attractiveness cues on JOLs in isolation. In daily life, the process of recognizing and memorizing a single face is often simultaneously influenced by both emotion and attractiveness. Moreover, when a face contains emotional information, its attractiveness level often changes as well. According to the reinforcement-affect theory of attraction, smiling faces appear more pleasant, meaning that happy facial expressions increase facial attractiveness, whereas negative expressions decrease it [?]. Therefore, emotion and attractiveness cues may mutually influence each other during facial metamemory monitoring. This process of multiple cues influencing metamemory monitoring involves cue integration.

Cue integration refers to the process in which individuals base their confidence ratings on two or more valid cues in non-single-cue situations [?, ?, ?]. Whether participants engage in cue integration can be determined by individually assessing whether each cue influences a given participant' s JOLs. There are two indicators: First, mean difference, calculated by testing for mean differences in JOL ratings across different levels of a cue for each participant—if the difference is significant, it indicates that the participant used that cue when making JOLs. Second, effect size, also called cue effect, calculated by obtaining the effect size from the mean difference test—if the effect size reaches the specified threshold for a small effect (even without significant mean difference), it suggests the participant' s JOLs were based on that cue. If a participant based their JOLs on two or more cues, this indicates cue integration occurred. However, the essence of effect size is the magnitude of mean difference; only when differences are significant is the effect size more meaningful [?]. Therefore, when using the effect size method for cue integration statistics, both significant mean differences and meeting the small effect size threshold should be required to demonstrate that individuals based their JOLs on a cue. However, existing studies using the effect size indicator for cue integration have not considered the prerequisite of significant mean differences [?]. Consequently, this study adopts a more stringent effect size indicator that ensures both significant mean differences and meeting the small effect size threshold to examine cue integration in the JOL process.

In reality, there are individual differences in whether participants integrate cues during JOLs in multi-cue situations. Previous studies have distinguished individuals who integrate cues from those who do not using the aforementioned mean difference or cue effect methods, then calculated the proportion of participants who engaged in cue integration [?, ?]. For example, word-pair learning studies found that 54% of participants made JOLs based on integration of font size and frequency cues, while 46% did not integrate cues—some relied on only one cue, and others did not use either of the two cues at all [?]. Similarly, another study found that most participants integrated word concreteness and emotion cues when making JOLs; the researchers further compared memory performance between cue integration and non-integration groups and found the former was higher, but did not compare JOL accuracy between the two groups [?]. Comparable results were found in studies using scene picture materials. Researchers manipulated two scene picture-related cues and found that 46.3% of participants integrated both cues when making JOLs, while the remaining participants did not engage in cue integration [?]. In summary, individual differences in cue integration during the JOL stage exist across different memory materials (verbal materials and scene pictures). Therefore, when examining how facial emotion and attractiveness cue integration affects JOL accuracy, this study must first distinguish between cue integration and non-integration groups using the data analysis methods from the aforementioned studies.

Existing research on cue integration in JOLs has only investigated behavioral aspects and has not revealed the physiological mechanisms underlying how cue integration affects JOL accuracy. ERPs studies have found that the frontal region is closely related to the JOL process. The word-list paradigm discovered that JOLs are associated with late positive waves in the frontal to central regions [?, ?]. Müller et al. (2016) found similar results, with task-related positive slow waves (PSW) appearing in the frontal region when participants made JOLs. Therefore, this study localized EEG components during the JOL stage to the frontal region. Additionally, previous studies have only examined whether EEG components from either the encoding or JOL stage alone are significantly correlated with JOLs. However, JOL accuracy involves neural activity across multiple stages. Moreover, when asked, participants reported engaging in the JOL process during both the late encoding stage before JOLs began and throughout the entire learning stage. Therefore, investigating EEG components related to JOL accuracy across multiple stages can provide deeper understanding of the temporal characteristics of confidence rating processes.

Furthermore, ERPs studies on face memory have found that the occipital P1 component and parietal late positive potential (LPP) are both related to face processing, while frontal negative-going slow waves (NSW) are related to face information maintenance. Specifically, P1 occurs in the early stage and primarily processes low-level facial physical features [?, ?, ?]. LPP occurs in the parietal region, typically appearing in the late stage of face encoding, and is related to emotion and attractiveness. It reflects high-level attentional resource allocation and cognitive evaluation, and is also associated with emotion regulation, marking emotional stimuli for encoding and consolidation [?, ?]. The frontal NSW component appears in the maintenance stage after memory encoding and is related to information maintenance and rehearsal [?, ?]. Face memory studies have also found that frontal negative-going slow waves (NSW) are related to information maintenance [?]. Previous studies have examined ERPs components related to face memory across multiple stages, and the process of JOL monitoring memory also involves multiple stages. Therefore, after adding JOLs, EEG components related to memory across various stages may be associated with JOLs and their accuracy.

This study aims to investigate how emotion and attractiveness cue integration affects JOL accuracy and its multi-stage temporal characteristics. The pilot experiment preliminarily examined the number of participants who integrated emotion and attractiveness cues during JOLs to ensure the effectiveness of the two-cue manipulation and guarantee balanced group sizes for subsequent experiments. Experiment 1 increased the intensity gradient of each attractiveness cue level based on pilot results, calculated the cue effect for each participant across the two cues, and accordingly divided participants into cue integration and non-integration groups to further explore whether cue integration affects memory and JOL accuracy. Experiment 2 incorporated ERPs technology to investigate the reasons why cue integration facilitates JOL accuracy. JOL accuracy is the correlation between memory performance and JOL ratings, so improved JOL accuracy may involve the encoding, JOL, and retrieval stages. Since participants knew in advance that they would make JOLs for each encoded item, they may have processed items for memory in the early encoding stage while already beginning the cognitive evaluation process for JOLs in the late encoding stage, continuing through to the JOL stage where information was rehearsed and maintained until the JOL response keypress ended. Therefore, this study used face pictures to explore the temporal characteristics of JOL accuracy, expecting to find LPP components related to high-level cognitive resource allocation in the late encoding stage and NSW components related to information maintenance and rehearsal in the JOL stage, both predicting JOL accuracy. In contrast, P1 primarily involves low-level early visual processing, with less involvement in deeper, higher-level cognitive processing, and is mainly related to processing stimulus physical features. Therefore, we expected that the P1 component would not be correlated with JOL accuracy.

2 Pilot Experiment: Investigating Cue Integration Levels of Emotion and Attractiveness in JOLs

This study aimed to investigate how cue integration affects JOL accuracy for face memory, selecting emotion and attractiveness as the two cues. The prerequisite for better comparing the effects of the two cues was ensuring consistency in manipulation methods, namely selecting a unified cue intensity variable: emotion intensity for the emotion cue and attractiveness intensity for the attractiveness cue. The pilot experiment' s purpose was to explore the number of participants who integrated versus did not integrate emotion and attractiveness cues, to determine whether the manipulation of each cue level was appropriate and to ensure balanced group sizes for subsequent experiments.

2.1 Method

2.1.1 Participants We used G*Power software to determine the sample size before the experiment [**?**, **?**, **?**]. With effect size defined as medium (0.25), α error probability as 0.05, and power as 0.8, the calculated required sample size was 13 participants.

We recruited 21 university students (9 male) from Northeast Normal University, all with normal or corrected-to-normal vision, right-handed, and without psychiatric or psychological disorders. Before the experiment, participants were informed of the experimental purpose, procedures, and precautions using standardized instructions, then signed informed consent forms and received compensation after completing the experiment. Due to 2 participants withdrawing midway, 19 valid participants (8 male) were included. The experimental protocol was approved by the Ethics Committee of the School of Psychology at Northeast Normal University.

2.1.2 Materials All experimental materials were selected from the Chinese Facial Affective Picture System [**?**]. We recruited 24 university students (11 male) from Northeast Normal University to rate 480 face pictures on attractiveness intensity and emotional intensity using a 0-100 scale. All participants' ratings were averaged, and pictures were sorted from high to low scores. The top 30% were selected as high intensity, the bottom 30% as low intensity, and the middle 40% as medium intensity. Pictures scoring above 50 were selected as high emotional intensity, 40-50 as medium emotional intensity, and below 40 as low emotional intensity. Pictures scoring above 45 were selected as high attractiveness intensity, 35-45 as medium attractiveness intensity, and below 35 as low attractiveness intensity.

Based on participants' rating results, nine types of face pictures were created: high emotion-high attractiveness, high emotion-medium attractiveness, high emotion-low attractiveness, medium emotion-high attractiveness, medium emotion-medium attractiveness, medium emotion-low attractiveness, low emotion-high attractiveness, low emotion-medium attractiveness, and low emotion-low attractiveness. We randomly selected 30 pictures of each type, balanced for valence (15 positive, 15 negative). Twenty pictures of each type were used in the learning phase (180 face pictures total in the learning phase), and the remaining 10 were used as distractors in the test phase (90 new pictures and 180 old pictures in the test phase). All face pictures were presented on a black background with identical brightness, and all responses were recorded.

2.1.3 Procedure The experiment consisted of learning and test phases, as shown in Figure 1 [Figure 1: see original paper]. In the learning phase, there were nine types of face pictures with 20 pictures each, totaling 180 trials. In each trial, a "+" was first presented at the center of the screen for 500 ms, followed by a face picture at the same location for 2000 ms. Participants were instructed to memorize the face picture. After the picture disappeared, participants were asked to provide a JOL rating by responding to the prompt "How confident are you that you will correctly recognize this photo in the subsequent test? Rate from 0-100." Participants were required to input a number from 0-100 using the keyboard based on their true confidence level, with higher numbers indicating higher confidence in future recognition. After entering the number, participants pressed "Enter" to proceed to the next trial.

After the learning phase, participants completed a 3-minute distraction task (continuously subtracting 3 from 1000), followed by the test phase. The test phase consisted of 270 trials, including 180 old pictures from the learning phase and 90 new pictures not previously presented, all in random order. In each trial, a fixation point was presented for 500 ms, followed by a picture for 4000 ms. Participants were required to judge within 4000 ms whether the picture was old (R key) or new (Y key) from the learning phase, pressing the corresponding key with their index finger. They then made an R/K judgment about their old/new decision: the screen displayed "1 = remember clearly, 2 = remember but not sure, 3 = guess," and participants responded according to their true experience. If they clearly remembered the picture was old/new, they pressed 1; if they remembered but were not sure, they pressed 2; if the judgment was pure guesswork, they pressed 3. After the keypress, the next trial began. Trials with R/K judgments of 3 were excluded from data analysis to eliminate guessing probability.

2.1.4 Design and Data Processing Data were statistically analyzed using SPSS 22.0. Single-subject analysis was conducted first, calculating cue effects for emotion and attractiveness cues for each participant. Mean JOL ratings were obtained for different levels of each cue, and one-way repeated-measures ANOVAs with three levels were conducted separately for each cue: emotional cue (high, medium, low intensity) and attractiveness cue (high, medium, low intensity). If differences were significant, it indicated that the participant showed a cue effect for that cue, meaning they used that cue when making JOLs; otherwise, the participant did not show that cue effect and did not use the cue for JOLs. For participants with significant differences, the effect size p^2 was calculated—the larger the p^2, the stronger the cue effect, indicating more extensive use of that cue in JOLs. Thus, some participants may have been exposed to the manipulated cues but did not actually use them when making JOLs. Based on whether each participant showed emotion and attractiveness cue effects, we tallied cue usage across all participants. Following previous research, participants showing both emotion and attractiveness cue effects were defined as the cue integration group, and the number of participants engaging in cue integration was counted.

2.2 Results

First, individual-level analysis was conducted: separate ANOVAs were performed on each participant' s JOL ratings across the three intensity levels for emotion and attractiveness cues, recording each participant' s p^2 for both cues. Using the criterion of p$^2 \geq 0.02$ [?], we recorded the number of cues each participant used in the JOL stage, thereby documenting their cue integration status. Specifically, 84% of participants based their confidence ratings on both emotion and attractiveness cues (engaging in two-cue integration), while 16% based their ratings only on the emotion cue, not using attractiveness as a basis for JOLs (see Table 1).

Supplementary analysis of cue effects was then conducted. Paired-sample t-tests were performed on emotion cue effect sizes and attractiveness cue effect sizes across all participants to compare whether the two cue effects differed significantly. Results showed $t(18) = 5.53$, $p < 0.001$, Cohen' s $d = 1.70$, with emotion cue effects ($M = 0.26 \pm 0.18$) significantly larger than attractiveness cue effects ($M = 0.04 \pm 0.03$), indicating that participants were less sensitive to attractiveness cues compared to emotion cues.

Table 1 Number (percentage) of participants basing JOLs on different cues

Cue Usage	Pre-Experiment	Experiment 1	Experiment 2
Using neither cue	0 (0%)	4 (10%)	5 (12%)
Using only emotion	3 (16%)	9 (22.5%)	11 (26%)
Using only attractiveness	0 (0%)	5 (12.5%)	3 (7%)
Using both cues (cue integration)	16 (84%)	22 (55%)	23 (55%)

Note: n = 19 for pre-experiment, n = 40 for Experiment 1, n = 42 for Experiment 2

2.3 Discussion

The pilot experiment investigated cue integration levels of facial emotion and attractiveness. We found that when both emotion and attractiveness cues were present, most participants (84%) tended to base their confidence judgments on both cues simultaneously—that is, they engaged in two-cue integration—though a minority (16%) based their JOLs only on the emotion cue. This is consistent with previous research showing that when multiple cues are available, most participants tend to base their JOLs on multiple cues [?]. Additionally, no participants in the pilot experiment based their JOLs solely on attractiveness. We speculate that when emotion and attractiveness cues coexist, participants are more sensitive to emotion cues—that is, they more easily notice emotional information in face pictures and consequently make JOLs based more heavily on differences in emotional intensity, which in turn affects memory and metamemory information maintenance and retrieval processes.

Supplementary analysis of cue effects revealed that the effect size for attractiveness cues was lower than that for emotion cues during the JOL stage, indicating lower cue effectiveness. This made participants less sensitive to attractiveness cues, resulting in zero participants using only attractiveness cues and creating a large disparity in group sizes between the cue integration and non-integration groups, making further analysis of cue integration effects on JOL accuracy difficult. Therefore, in Experiment 1, we increased the gradient between the three levels of attractiveness cues to improve their effectiveness for participants, thereby ensuring more similar group sizes between cue integration and non-integration participants as a basis for investigating how using versus not using both cues affects memory and JOL accuracy.

3 Experiment 1: Effects of Emotion and Attractiveness Cue Integration on JOL Accuracy

The pilot experiment found that participants were less sensitive to attractiveness cues than emotion cues, resulting in large differences between cue integration and non-integration group sizes. Therefore, Experiment 1 increased the gradient between the three levels of attractiveness cues to improve their effectiveness and ensure comparable group sizes, aiming to investigate whether emotion and attractiveness cue integration could promote memory and JOL accuracy.

3.1 Method

3.1.1 Participants We used G*Power software for power analysis to determine sample size before the experiment. With effect size defined as medium (0.25), α error probability as 0.05, and power as 0.8, the calculated required sample size was 13 participants.

We recruited 41 university students from Northeast Normal University, all right-handed with normal or corrected-to-normal vision and without psychiatric or psychological disorders. Before the experiment, participants were informed of the experimental purpose, procedures, and precautions using standardized instructions, then signed informed consent forms and received compensation after completing the experiment. Ultimately, 40 valid datasets (18 male) were retained. The experimental protocol was approved by the Ethics Committee of the School of Psychology at Northeast Normal University.

3.1.2 Materials All experimental materials were selected from the Chinese Facial Affective Picture System (CAFPS) based on pilot experiment rating results. The emotional intensity gradient of selected face pictures remained consistent with Experiment 1: above 50 as high emotional intensity, 40-50 as medium emotional intensity, and below 40 as low emotional intensity. Compared to the pilot experiment, Experiment 1 increased the attractiveness intensity gradient: the top 25% as high attractiveness intensity, the bottom 25% as low attractiveness intensity, and the middle 25% as medium attractiveness intensity. Thus, pictures scoring above 60 were defined as high attractiveness intensity, 40-50 as medium attractiveness intensity, and below 30 as low attractiveness intensity.

We randomly selected nine types of face pictures: high emotion-high attractiveness, high emotion-medium attractiveness, high emotion-low attractiveness, medium emotion-high attractiveness, medium emotion-medium attractiveness, medium emotion-low attractiveness, low emotion-high attractiveness, low emotion-medium attractiveness, and low emotion-low attractiveness. Each type included 30 pictures with balanced valence (15 positive, 15 negative). Twenty pictures of each type were used in the learning phase (180 pictures total), and the remaining 10 were used as distractors in the test phase (90 new pictures and 180 old pictures in the test phase). Background presentation was identical to the pilot experiment, and all responses were recorded.

3.1.3 Procedure The procedure was identical to the pilot experiment.

3.1.4 Design and Data Processing The experiment used a 2 (group: cue integration vs. non-integration) × 3 (emotion cue: high, medium, low intensity) × 3 (attractiveness cue: high, medium, low intensity) mixed design, with group as a between-subjects variable (based on cue effects in the JOL stage) and emotion and attractiveness cues as within-subjects variables. Data were analyzed using repeated-measures ANOVA in SPSS 22.0. Dependent variables included: hit rate, JOL ratings, Gamma correlation between memory performance and JOLs, and JOL effect size p^2 (cue effect).

The data analysis approach was consistent with the pilot experiment, with additional analysis of cue integration effects on memory and JOLs—that is, examining differences between cue integration and non-integration groups. Therefore, after calculating cue effects for all participants and documenting cue integration status, participants were divided into cue integration and non-integration groups. Group (cue integration vs. non-integration) was then included as a between-subjects independent variable in a mixed-design ANOVA with emotion and attractiveness cues to investigate how cue integration affects memory and JOL accuracy.

3.2 Results

3.2.1 Determining Cue Integration and Non-Integration Group Sizes Individual-level analysis was conducted: each participant's p^2 was recorded, and using the criterion $p^2 \geq 0.02$, cue integration status was documented. Specifically, 22 participants (55%) based their confidence ratings on both emotion and attractiveness cues when making JOLs; 9 participants (22.5%) based their ratings only on emotion cues, not using attractiveness; 5 participants (12.5%) based their ratings only on attractiveness cues, not using emotion; and 4 participants (10%) did not base their ratings on either cue. Participants who used both cues were classified as the cue integration group, and the rest as the non-integration group (see Table 1).

3.2.2 Effects of Cue Integration on Memory and JOL Accuracy Participants who based their JOLs on both emotion and attractiveness cues were assigned to the cue integration group, while those who used only attractiveness cues were assigned to the non-integration group. Group was included as a between-subjects independent variable in a 2 (group: cue integration vs. non-integration) × 3 (emotion cue: high, medium, low intensity) × 3 (attractiveness cue: high, medium, low intensity) repeated-measures ANOVA. Descriptive statistics are shown in Table 2 .

A repeated-measures ANOVA on hit rates focused only on results related to cue integration. The analysis revealed a significant interaction between group and attractiveness, $F(2,76) = 4.64$, $p = 0.038$, $p^2 = 0.109$. For medium and low attractiveness faces, the cue integration group showed significantly higher hit rates ($M = 0.55 \pm 0.04$; $M = 0.46 \pm 0.04$) than the non-integration group ($M = 0.52 \pm 0.04$; $M = 0.46 \pm 0.04$). No significant differences were found between groups for high attractiveness faces. The main effect of group was not significant, $F(1,38) = 1.21$, $p > 0.05$, $p^2 = 0.031$. The interaction between group and emotion was not significant, $F(2,76) = 1.80$, $p > 0.05$, $p^2 = 0.045$. The three-way interaction between group, emotion, and attractiveness was not significant, $F(4,152) = 1.70$, $p > 0.05$, $p^2 = 0.043$.

A repeated-measures ANOVA on JOL accuracy focused only on results related to cue integration. The analysis revealed a significant main effect of group, $F(1,38) = 14.84$, $p < 0.001$, $p^2 = 0.281$. The cue integration group showed significantly higher JOL accuracy ($M = 0.27 \pm 0.03$) than the non-integration group ($M = 0.10 \pm 0.03$). The interaction between group and attractiveness was not significant, $F(2,76) = 0.89$, $p > 0.05$, $p^2 = 0.023$. The interaction between group and emotion was not significant, $F(2,76) = 0.15$, $p > 0.05$, $p^2 = 0.004$. The three-way interaction between group, emotion, and attractiveness was not significant, $F(4,152) = 2.37$, $p > 0.05$, $p^2 = 0.059$.

Table 2 Descriptive statistics for cue integration and non-integration groups ($n_1 = 22$, $n_2 = 18$)

Condition	Cue Integration Group	Non-Integration Group
Hit Rate	M ± SD	M ± SD
High Emotion-High Attractiveness	0.74 ± 0.18	0.76 ± 0.17
High Emotion-Medium Attractiveness	0.63 ± 0.23	0.68 ± 0.25
High Emotion-Low Attractiveness	0.63 ± 0.23	0.58 ± 0.31
Medium Emotion-High Attractiveness	0.52 ± 0.24	0.59 ± 0.28
Medium Emotion-Medium Attractiveness	0.55 ± 0.04	0.52 ± 0.04
Medium Emotion-Low Attractiveness	0.46 ± 0.04	0.46 ± 0.04
Low Emotion-High Attractiveness	0.45 ± 0.24	0.48 ± 0.25
Low Emotion-Medium Attractiveness	0.42 ± 0.22	0.45 ± 0.23
Low Emotion-Low Attractiveness	0.38 ± 0.21	0.40 ± 0.22
Gamma Correlation	0.27 ± 0.03	0.10 ± 0.03

Note: n_1 = number in cue integration group, n_2 = number in non-integration group; H = high intensity, M = medium intensity, L = low intensity; first letter represents emotion cue, second represents attractiveness cue (e.g., HM = high emotion-medium attractiveness); data presented as M ± SD

3.3 Discussion

Analysis based on cue integration statistics revealed that over half of the participants simultaneously based their JOLs on both emotion and attractiveness cues—that is, they engaged in cue integration. This is consistent with the pilot experiment results and previous research [?]. Analysis of cue integration effects on memory and JOL accuracy found that the main effect of group on hit rate was not significant, meaning that hit rates did not differ statistically between the cue integration and non-integration groups. Previous studies using mean difference as the cue integration indicator found that cue integration could improve memory performance—that is, significant differences in hit rates existed between cue integration and non-integration groups [?]. However, when using cue effect (effect size) as the cue integration indicator, no differences in memory performance were found between integration and non-integration groups, consistent with our results. Our study' s more stringent method of defining cue integration using the effect size indicator made participant classification in cue integration more precise, consequently making differences in hit rates less likely to emerge. Notably, the main effect of group on JOL accuracy was significant, with the cue integration group showing significantly higher JOL accuracy than the non-integration group. This suggests that participants who integrated multiple cues incorporated more information from the learning materials into their confidence assessments, enabling better monitoring of their own memory cognitive processes for face learning, which in turn allowed them to adjust their learning states and cognitive resources to improve JOL accuracy. Experiment 2 further employed ERPs technology to investigate EEG waveforms across multiple stages to seek the reasons why cue integration facilitates face memory JOL accuracy.

4 Experiment 2: Temporal Characteristics of Cue Integration Effects on JOL Accuracy

Experiment 1 found that the cue integration group showed significantly higher JOL accuracy than the non-integration group, indicating that simultaneously using both facial emotion and attractiveness cues was more effective for integrating facial information and improving JOL accuracy than using only one cue or neither cue. To reveal the reasons for this phenomenon, Experiment 2 used ERPs technology to further explore the multi-stage temporal characteristics of how cue integration facilitates face picture JOL accuracy. To increase participants' sensitivity to both cues and facilitate observation of EEG amplitude differences, we further increased the intensity level gradients for both emotion and attractiveness cues, changing from three levels to two levels to investigate whether the cue integration group would show larger amplitude of cue-related components than the non-integration group during encoding, JOL, and retrieval stages.

4.1 Method

4.1.1 Participants We used G*Power software for power analysis to determine sample size before the experiment. With effect size defined as medium (0.25), α error probability as 0.05, and power as 0.8, the calculated required sample size was 24 participants.

We recruited 45 university students from Northeast Normal University with the same requirements as Experiment 1. Ultimately, 42 valid datasets (20 male) were retained.

4.1.2 Materials All experimental materials were selected from the Chinese Facial Affective Picture System (CAFPS) based on pilot experiment rating results. Pictures scoring above 60 were selected as high emotional intensity, and below 30 as low emotional intensity. Pictures scoring above 60 were selected as high attractiveness intensity, and below 30 as low attractiveness intensity. We randomly selected four types of face pictures: high emotion-high attractiveness, high emotion-low attractiveness, low emotion-high attractiveness, and low emotion-low attractiveness. Each type included 75 pictures with balanced valence (37 positive, 38 negative). Fifty pictures of each type were used in the learning phase (200 pictures total), and the remaining 25 were used as distractors in the test phase (100 new pictures and 200 old pictures in the test phase). Background presentation was identical to the pilot experiment, and all responses were recorded.

4.1.3 Procedure To ensure valid EEG data, compared to Experiment 1, Experiment 2 changed the JOL stage from a 0-100 keypress to two keypress options: high confidence and low confidence. Additionally, to ensure adequate EEG component collection, a 500 ms blank screen was added after both the stimulus and response screens. The rest of the procedure remained consistent with Experiment 1. Before the formal experiment, participants were instructed to minimize blinking and head movements, especially during stimulus presentation and response. See Figure 2 [Figure 2: see original paper].

4.1.4 EEG Recording The experimental program was written in E-Prime 3.0 and displayed on a Lenovo 17-inch monitor (1024$×$768). All participants sat in front of the computer with eyes approximately 90 cm from the display. We used a Neuroscan SynAmps2 amplifier to acquire online EEG data and a Neuroscan Quik-Cap 64 electrode cap, recording EEG from 62 electrode sites according to the international standard 10-20 extended system. The left mastoid (M1) served as the online reference. Horizontal electrooculogram (HEOG) was recorded from electrodes placed approximately 1 cm lateral to the outer canthus of each eye, and vertical electrooculogram (VEOG) from electrodes placed approximately 1.5 cm above and below the left eye. All electrode impedances were maintained below 10 kΩ throughout testing. The sampling rate was 1000 Hz with a filter bandwidth of 0.05-100 Hz.

4.1.5 Design and Data Analysis The experiment used a 2 (group: cue integration vs. non-integration) × 2 (emotion cue: high, low intensity) × 2 (attractiveness cue: high, low intensity) mixed design, with group as a between-subjects variable (based on cue effects in JOLs) and emotion and attractiveness cues as within-subjects variables. ANOVA was conducted using SPSS 22.0. Dependent variables were identical to Experiment 2: hit rate, JOL ratings, Gamma correlation between memory performance and JOLs, and JOL effect size ϕ (cue effect).

EEG data were analyzed using the EEGLAB toolbox [?] in Matlab 2018. Data were first down-sampled to 500 Hz and re-referenced to the average of bilateral mastoids (M1 + M2)/2. A high-pass filter of 0.1 Hz and low-pass filter of 40 Hz were applied [?]. Independent component analysis (ICA) was then performed to remove artifacts related to blinking. Trials with absolute EEG voltage exceeding $±$100 µV were excluded. Participants with more than 35% of trials excluded were removed from final analysis, resulting in 3 participants being excluded. Ultimately, 42 participants provided valid data.

Based on our research objectives, ERPs evoked after stimulus presentation in the encoding, JOL, and retrieval stages were analyzed. The 200 ms before stimulus presentation served as baseline, and the epoch from baseline to 1000 ms after stimulus presentation was used for EEG analysis, with segmentation according to experimental conditions. All three stages could be divided into high emotion-high attractiveness, high emotion-low attractiveness, low emotion-high attractiveness, and low emotion-low attractiveness conditions; in the retrieval stage, only hit trials were retained.

This study focused on NSW, LPP, and P1 components. Based on previous research, LPP was measured as the average of CP1, CPZ, and CP2 electrodes in the 650-680 ms time window; NSW was measured as the average of F1, FZ, and F2 electrodes in the 760-860 ms time window; and P1 was measured as the average of O1, OZ, and O2 electrodes in the 90-120 ms time window [?, ?, ?]. Matlab software was used for statistical analysis of amplitude in each time window, conducting independent samples t-tests (group: cue integration vs. non-integration).

4.2 Results

4.2.1 Effects of Cue Integration on Memory and JOL Accuracy Individual-level analysis was conducted: ^{2} tests were performed on each participant' s JOLs for each cue to obtain effect size ϕ for each participant and cue. Using the criterion $\phi \geq 0.2$ [?], cue integration status was recorded. Specifically, 23 participants (55%) based their confidence ratings on both emotion and attractiveness cues when making JOLs, 11 participants (26%) used only emotion cues, 3 participants (7%) used only attractiveness cues, and 5 participants (12%) did not use either cue (see Table 1). Participants who used both cues were classified as the cue integration group, and the rest as the non-integration group.

Participants who used both emotion and attractiveness cues were assigned to the cue integration group, while those who used only attractiveness cues and those who used neither cue were assigned to the non-integration group. Group was included as a between-subjects independent variable in a 2 (group: cue integration vs. non-integration) × 2 (emotion cue: high, low intensity) × 2 (attractiveness cue: high, low intensity) mixed-design ANOVA. Descriptive statistics are shown in Table 3 .

A repeated-measures ANOVA on hit rates focused only on results related to cue integration. The analysis revealed no significant main effect of group, $F(1,40) = 0.17$, $p > 0.05$, $p^2 = 0.004$. The interaction between group and emotion was not significant, $F(1,40) = 0.12$, $p > 0.05$, $p^2 = 0.003$. The interaction between group and attractiveness was not significant, $F(1,40) = 1.48$, $p > 0.05$, $p^2 = 0.036$. The three-way interaction between group, emotion, and attractiveness was not significant, $F(1,40) = 1.17$, $p > 0.05$, $p^2 = 0.028$.

A repeated-measures ANOVA on JOL accuracy focused only on results related to cue integration. The analysis revealed a significant main effect of group, $F(1,40) = 8.01$, $p = 0.007$, $p^2 = 0.17$. The cue integration group showed significantly higher JOL accuracy ($M = 0.49 \pm 0.05$) than the non-integration group ($M = 0.29 \pm 0.05$). The interaction between group and emotion was not significant, $F(1,40) = 0.24$, $p > 0.05$, $p^2 = 0.006$. The interaction between group and attractiveness was not significant, $F(1,40) = 0.22$, $p > 0.05$, $p^2 = 0.005$. The three-way interaction between group, emotion, and attractiveness was not significant, $F(1,40) = 1.45$, $p > 0.05$, $p^2 = 0.035$.

Table 3 Descriptive statistics for cue integration and non-integration groups ($n_1 = 23$, $n_2 = 19$)

Condition	Cue Integration Group	Non-Integration Group
Hit Rate	M ± SD	M ± SD
High Emotion-High Attractiveness	0.74 ± 0.18	0.76 ± 0.17
High Emotion-Low Attractiveness	0.63 ± 0.23	0.68 ± 0.25
Low Emotion-High Attractiveness	0.63 ± 0.23	0.58 ± 0.31
Low Emotion-Low Attractiveness	0.52 ± 0.24	0.59 ± 0.28
Gamma Correlation	0.49 ± 0.05	0.29 ± 0.05

Note: n_1 = number in cue integration group, n_2 = number in non-integration group; data presented as M ± SD

Table 4 ERPs amplitude statistics for cue integration and non-integration groups ($n_1 = 23$, $n_2 = 19$)

Component	Cue Integration Group	Non-Integration Group
LPP (V)	6.15 ± 4.47	3.41 ± 4.07
NSW (V)	-7.89 ± 3.98	-4.53 ± 5.24
P1 (V)	4.90 ± 3.95	7.28 ± 4.68

Note: n_1 = number in cue integration group, n_2 = number in non-integration group; data presented as M ± SD

4.2.2 ERPs Amplitude Comparison Between Cue Integration and Non-Integration Groups Based on the significant behavioral group main effect on JOL accuracy, we further investigated the neural mechanisms by conducting independent samples t-tests between groups for the encoding, JOL, and retrieval stages.

First, in the 650-680 ms time window after stimulus presentation during the encoding stage, we examined LPP amplitude differences between cue integration and non-integration groups at parietal electrodes CP1, CPZ, and CP2. Independent samples t-test analysis revealed a significant group main effect, $t(40) = 2.06$, $p = 0.046$, Cohen' s $d = 0.65$. The cue integration group ($M = 6.15 \pm 4.47$) showed larger LPP amplitude than the non-integration group ($M = 3.41 \pm 4.07$). Waveforms and topographic maps are shown in Figure 3 [Figure 3: see original paper], with descriptive amplitude statistics in Table 4.

In the 760-860 ms time window after stimulus presentation during the JOL stage, we examined NSW amplitude differences between groups at frontal electrodes F1, FZ, and F2. Independent samples t-test analysis revealed a significant group main effect, $t(40) = 2.36$, $p = 0.023$, Cohen's $d = 0.75$. The cue integration group ($M = -7.89 \pm 3.98$) showed larger NSW amplitude than the non-integration group ($M = -4.53 \pm 5.24$). Waveforms and topographic maps are shown in Figure 4 [Figure 4: see original paper], with descriptive amplitude statistics in Table 4.

In the 90-120 ms time window after stimulus presentation during the retrieval stage, we examined P1 amplitude differences between groups at occipital electrodes O1, OZ, and O2. Independent samples t-test analysis revealed no significant group main effect, $t(40) = 1.86$, $p > 0.05$, Cohen' s $d = 0.59$. Waveforms and topographic maps are shown in Figure 5 [Figure 5: see original paper], with descriptive amplitude statistics in Table 4.

Table 5 Pearson correlations between ERPs amplitudes and Gamma values ($n_1 = 23$, $n_2 = 19$)

Component	Cue Integration Group	Non-Integration Group
LPP	$r = 0.48$, $p = 0.019$	$r = 0.19$, $p = 0.43$
NSW	$r = 0.57$, $p = 0.004$	$r = 0.37$, $p = 0.12$
P1	$r = 0.41$, $p = 0.053$	$r = 0.03$, $p = 0.90$

Note: n_1 = number in cue integration group, n_2 = number in non-integration group

4.2.3 Correlation Between JOL Accuracy and ERPs Amplitudes Since JOL accuracy involves multiple stages—encoding, JOL, and retrieval—the brain activity in each stage may be associated with this behavioral performance. Therefore, to further demonstrate the relationship between cue integration and JOL accuracy, this study examined the connection between ERPs amplitudes and behavioral performance. Pearson correlations analyzed the relationships between LPP, NSW, and P1 amplitudes and JOL accuracy in both cue integration and non-integration groups. As shown in Table 5, results indicated that in the cue integration group, JOL accuracy showed significant positive correlation with frontal NSW amplitude, $r = 0.57$, $p = 0.004$, and with parietal LPP amplitude, $r = 0.48$, $p = 0.019$, but no significant correlation with occipital P1 amplitude, $r = 0.41$, $p > 0.05$. In the non-integration group, JOL accuracy showed no significant correlation with parietal LPP, frontal NSW, or occipital P1 amplitudes, $r = 0.19, 0.37, 0.03$, $p = 0.43, 0.12, 0.90$, respectively. These ERPs results, consistent with the behavioral findings, indicate that compared to the non-integration group, individuals in the cue integration group engaged in deeper cognitive evaluation of face pictures and allocated more cognitive resources to picture maintenance and rehearsal during the JOL stage, evoking more positive amplitudes that positively correlated with behavioral responses. In contrast, the non-integration group did not engage in more detailed memory and evaluation based on the manipulated cues, resulting in lower JOL accuracy.

4.3 Discussion

The cue integration statistics in Experiment 2 were similar to those in the pilot experiment and Experiment 1, with over half of participants integrating both facial emotion and attractiveness cues, consistent with previous research. In the analysis of cue integration effects on memory and metamemory, we found that cue integration improved JOL accuracy but not memory performance, consistent with Experiment 1 results. Because each cue had only two levels in Experiment 2, participants' cognitive load was lower than in Experiment 1, so regardless of whether they integrated the two cues, participants achieved relatively good performance, making group differences less likely to emerge.

In the ERPs results, the late positive component (LPP) related to emotional cognitive evaluation in the parietal region during the encoding stage was associated with cue integration, with the cue integration group showing significantly larger LPP amplitude than the non-integration group. This indicates that the cue integration group engaged in higher-level cognitive evaluation of face pictures compared to the non-integration group. In the frontal region during the JOL stage, the late negative slow wave (NSW) related to memory maintenance was associated with cue integration, with the cue integration group showing larger NSW amplitude than the non-integration group, indicating deeper rehearsal of information during the JOL stage. Moreover, both components showed significant positive correlations with the behavioral JOL accuracy measure, suggesting that the neural mechanisms underlying cue integration' s enhancement of metamemory monitoring accuracy are manifested in frontal late negative waves and parietal late positive waves.

5 General Discussion

This study used the classic study-JOL-recall paradigm across three experiments with face materials to investigate emotion and attractiveness cue integration and whether cue integration improves JOL accuracy and its multi-stage temporal characteristics. The pilot experiment manipulated high, medium, and low levels for both emotion and attractiveness cues, primarily exploring cue integration levels for face pictures. Results showed that 84% of participants engaged in two-cue integration—that is, they based their JOL ratings on both emotion and attractiveness cues—while the remaining participants used only emotion cues, indicating that attractiveness cues were not effective for this subset. Experiment 1 further increased the intensity gradient between attractiveness cue levels to enhance its effectiveness and explore whether two-cue integration affects memory and JOL accuracy. Results showed that 55% of participants engaged in cue integration, while the rest did not. Additionally, the cue integration group showed significantly improved JOL accuracy compared to the non-integration group, suggesting that when participants based their JOLs on multiple cues, they better monitored their cognitive processes for memorizing learning materials, enabling them to select better learning strategies and improve metamemory monitoring accuracy. In Experiment 2, we further increased the intensity differences between emotion and attractiveness cue levels and changed each cue to high and low two levels to investigate the EEG temporal characteristics of two-cue integration effects on memory and metamemory. Results showed that the late positive component (LPP) in the parietal region during the encoding stage and the late negative component (NSW) in the frontal region during the JOL stage were significantly correlated with the metamemory monitoring accuracy behavioral measure. Moreover, the cue integration group showed larger LPP amplitude than the non-integration group, indicating deeper cognitive evaluation during the encoding stage. The cue integration group also showed larger NSW amplitude than the non-integration group, indicating better maintenance and rehearsal of picture materials during the JOL stage, enabling more accurate JOL processes based on both cues.

Across the pilot experiment and two formal experiments, we consistently found that most participants engaged in two-cue integration of emotion and attractiveness. This is consistent with previous research. Studies using verbal materials found that over half of participants integrated font size and frequency cues [?]. Studies on scene pictures investigating JOL cue integration in Experiments 1 and 2 also found that over half of participants engaged in two-cue integration [?]. When multiple cues are manipulated, each manipulated cue becomes a basis for participants' JOLs. However, for participants who used all cues, the effectiveness of each cue decreased. This pattern differs significantly from JOL studies manipulating single cues. According to Kahneman' s (1973) cognitive resource limitation theory proposed in *Attention and Effort*, participants who engaged in cue integration simultaneously attended to both emotion and attractiveness cues during the JOL stage, consuming more attentional resources overall and consequently reducing utilization of individual cues, as reflected in decreased cue effects and less thorough use of each cue' s information compared to participants who relied on single cues.

Emotion and attractiveness cue integration facilitates JOL accuracy. In both Experiments 1 and 2, participants who integrated emotion and attractiveness cues during the JOL stage showed significantly higher JOL accuracy than those who used only single cues or neither cue. Some interactive effects and correlations between emotion and attractiveness cues enabled participants to integrate more learning material information, leading to reinforcement that allowed them to adjust their learning states and cognitive resources, conduct more accurate confidence assessments, and improve metamemory monitoring accuracy [?]. Additionally, because each cue had only two levels in Experiment 2, there was a risk of demand characteristics threatening cue effectiveness for JOLs—that is, when single cues are manipulated at two easily distinguishable levels, cue effects on JOLs might be based on participants' assumptions about what memory patterns would make the experiment more successful [?]. However, JOL accuracy results showed that cues remained effective for the JOL process, largely ruling out demand characteristics as an explanation.

The improvement in JOL accuracy from cue integration was manifested in deeper cognitive evaluation during the late encoding stage and information maintenance and rehearsal during the JOL stage. In Experiment 2, the late positive component (LPP) related to emotional cognitive evaluation in the parietal region during the encoding stage was associated with cue integration, with the cue integration group showing significantly larger LPP amplitude than the non-integration group. LPP occurs in the parietal region and is related to high-level cognitive resource allocation. For example, research has shown that emotional stimuli evoke higher LPP amplitude than neutral stimuli to obtain more attentional resources [?]. A study on emotion regulation found that increased LPP amplitude reflects increased attentional and other cognitive resources and deeper processing of motivational stimuli, indicating that the cue integration group allocated more cognitive resources and evaluation to face pictures than the non-integration group [?]. The late negative slow wave (NSW) in the frontal region during the JOL stage was associated with cue integration, with the cue integration group showing larger NSW amplitude than the non-integration group. Larger NSW amplitude in the frontal region indicates deeper information maintenance and rehearsal [?]. Both human and non-human primate prefrontal cortex regions have shown negative slow waves related to visual information maintenance [?], indicating that cue integration engaged in deeper rehearsal of information during the JOL stage. Moreover, both components showed significant positive correlations with the behavioral JOL accuracy measure, consistent with previous research finding late negative components in the frontal region related to JOL accuracy [?].

This study is the first to adopt a more stringent effect size indicator for cue integration. Previous studies investigating cue integration in JOLs for verbal and scene picture materials found that most participants based their JOLs on multiple cues and that cue integration improved memory performance to some extent [?, ?, ?]. However, when using effect size as the dependent variable for cue integration, these studies did not consider the prerequisite of significant mean differences between cue levels, potentially compromising the indicator’ s rigor. Therefore, this study improved the effect size indicator for cue integration and used face pictures to investigate emotion and attractiveness cue integration in JOLs, finding similar results—most participants simultaneously based their JOLs on both emotion and attractiveness cues. Moreover, when we artificially increased the differences between attractiveness cue levels, the effectiveness of the attractiveness cue improved, leading more participants to base their JOLs solely on that cue. The cue utilization model categorizes cues affecting metamemory into different types but does not consider interactions between cues or between cues and participants. Therefore, this study’ s results enrich the cue utilization model to some extent, further confirming that not all cues are effective for all participants and proposing methods to increase cue effectiveness.

Finally, this study has several limitations. First, it only demonstrates cue integration for emotion and attractiveness, both of which are internal cues. Whether different results would emerge with different cue types remains unknown, limiting generalization to integration of more than two cues or other cue types. Future research should investigate integration of more numerous cues in face materials. Second, metamemory monitoring measures were limited. This study focused on JOLs and their accuracy in face learning tasks and did not include other metamemory monitoring indicators. However, other prospective metamemory monitoring measures and retrospective metamemory monitoring are also important for memory processes [?, ?]. Future research could compare different metamemory monitoring types in the cue integration domain. Third, the participant population lacked diversity. This study used university students whose cognitive functions were fully developed and at their peak. As JOLs represent a high-level cognitive activity, functional differences may exist across age groups [?, ?]. Investigating behavioral and electrophysiological mechanisms of cue integration effects on JOL accuracy in other populations is also necessary to provide insights into the longitudinal development of metamemory monitoring accuracy.

This study conducted a pilot experiment and two formal experiments to explore whether emotion and attractiveness cue integration for face materials could improve JOL accuracy and its multi-stage temporal characteristics. The following conclusions were reached:

1. In face memory, over half of participants tend to simultaneously base their JOLs on both emotion and attractiveness cues—that is, they engage in emotion and attractiveness cue integration.
2. Cue integration facilitates JOL accuracy, manifested as deeper cognitive evaluation during the late encoding stage and better information maintenance and rehearsal during the late JOL stage.

References

Bröder, A., & Undorf, M. (2019). Metamemory viewed through the judgment lens. *Acta Psychologica, 19*(7), 153-165.

Byrne, D., & Clore, G. L. (1970). A reinforcement model of evaluative responses. *Personality: An international journal.*

Cauvin, S., Moulin, C. J. A., Souchay, C., Kliegel, M., & Schnitzspahn, K. M. (2019). Prospective memory predictions in aging: Increased overconfidence in older adults. *Experimental Aging Research, 45*(5), 436-459.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.*

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*(1), 9-21.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses.* Behavior Research Methods, 41*(4), 1149–1160.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences.* Behavior Research Methods, 39*(2), 175–191.

Fields, E. C. (2023). The P300, the LPP, context updating, and memory: What is the functional significance of the emotion-related late positive potential?. *International Journal of Psychophysiology.*

Flurry, E. (2021). Facilitating multiple cue use eliminates age differences in episodic metamemory for emotional words.

Gong, X., Huang, Y. X., Wang, Y., & Luo, Y. J. (2011). Revision of the Chinese facial affective picture system. *Chinese mental health journal.*

Hajcak, G., Dunning, J. P., & Foti, D. (2009). Motivated and controlled attention to emotion: Time-course of the late positive potential. *Clinical Neurophysiology, 120*(3), 505-510.

Hourihan, K. L., & Bursey, E. (2017). A misleading feeling of happiness: Metamemory for positive emotional and neutral pictures. *Memory (Hove, England), 25*, 35-43.

Hughes, G. I., & Thomas, A. K. (2022). When memory and metamemory align: How processes at encoding influence delayed judgment-of-learning accuracy. *Journal of Intelligence, 10*(4), 101.

Kahneman, D. (1973). *Attention and effort* (Vol. 1063, pp. 218-226). Englewood Cliffs, NJ: Prentice-Hall.

Kao, Y. C., Davis, E. S., & Gabrieli, J. (2005). Neural correlates of actual and predicted memory formation. *Nature Neuroscience, 8*(12), 1776-1783.

Koriat, A. (1997). Monitoring one' s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349-370.

Koriat, A., Ackerman, R., Lockl, K., & Schneider, W. (2009). The memorizing effort heuristic in judgments of learning: A developmental perspective. *Journal of Experimental Child Psychology, 102*(3), 265-279.

Legrand, N., Engen, S. S., Correa, C. M. C., Mathiasen, N. K., Nikolova, N., Fardo, F., & Allen, M. (2021). Emotional metacognition: stimulus valence modulates cardiac arousal and metamemory. *Cognition and Emotion, 35*(4), 705-721.

Liu, Y., Su, Y., Xu, G., & Pei, M. (2018). When do you know what you know? The emergence of memory monitoring. *Journal of Experimental Child Psychology, 166*, 34-48.

Loprinzi, P., Rigdon, B., Javadi, A. H., & Kelemen, W. L. (2022). Effects of acute exercise intensity on source episodic memory and metamemory accuracy. *Quarterly Journal of Experimental Psychology, 75*(12), 2232-2243.

Luck, S. J. (2014). *An introduction to the event-related potential technique.* MIT press.

Mcdonough, M., Baillie, A. J., Clark, P. J., Gowing, L., Stapinski, L. A., & Taye, B. (2021). Understanding and managing comorbidities for people with alcohol problems: Polydrug use and dependence, co-occurring mental disorders, and physical comorbidities. *The Medical Journal of Australia, (7), 215.*

Michaela, S., A, Matthias, D., Eva, L., & Ulrike, W. (2020). How influenceable is our metamemory for pictorial material? The impact of framing and emotionality on metamemory judgments. *Cognition, 195.*

Müller, B. C., Tsalas, N. R., van Schie, H. T., Meinhardt, J., Proust, J., Sodian, B., & Paulus, M. (2016). Neural correlates of judgments of learning—An ERP study on metacognition. *Brain Research, 16*(52), 170-177.

Mueller, M. L., & Dunlosky, J. (2017). How beliefs can impact judgments of learning: Evaluating analytic processing theory with beliefs about fluency. *Journal of Memory and Language, 93*, 245-258.

Reinhart, R. M., Heitz, R. P., Purcell, B. A., Weigand, P. K., Schall, J. D., & Woodman, G. F. (2012). Homologous mechanisms of visuospatial working memory maintenance in macaque and human: Properties and sources. *Journal of Neuroscience, 32*(22), 7711-7722.

Skavhaug, I. M., Wilding, E. L., & Donaldson, D. I. (2010). Judgments of learning do not reduce to memory encoding operations: Event-related potential evidence for distinct metacognitive processes. *Brain Research, 13*(18), 87-95.

Skavhaug, I. M., Wilding, E. L., & Donaldson, D. I. (2013). Immediate judgments of learning predict subsequent recollection: Evidence from event-related potentials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(1), 29-38.

Tatz, J. R., & Peynircioğlu, Z. F. (2019). Judgments of learning in context: Backgrounds can both reduce and produce metamemory illusions. *Memory and Cognition, 48*, 581-595.

Undorf, M., Amaefule, C. O., & Kamp, S. M. (2020). The neurocognitive basis of metamemory: Using the N400 to study the contribution of fluency to judgments of learning. *Neurobiology of Learning and Memory, 169*, 107176.

Undorf, M., & Bröder, A. (2021). Metamemory for pictures of naturalistic scenes: Assessment of accuracy and cue utilization. *Memory and Cognition, 49*(7), 1405-1422.

Undorf, M., Söllner, A., & Bröder, A. (2018). Simultaneous utilization of multiple cues in judgments of learning. *Memory and Cognition, 46*(4), 507-519.

Vaccaro, A. G., & Fleming, S. M. (2018). Thinking about thinking: A coordinate-based meta-analysis of neuroimaging studies of metacognitive judgements. *Brain and Neuroscience Advances, 2*, 1–14.

Wang, L., Gui, P., Li, L., Ku, Y., Bodner, M., Fan, G., & Dong, X. (2016). Neural correlates of heat-evoked pain memory in humans. *Journal of Neurophysiology, 115*(3), 1596-1604.

Yuchi, Z., & Li, P. (2014). Effect of face attractiveness on face recognition and judgments of learning. *Science of Social Psychology.*

Zhang, D. D., He, W. Q., Wang, T., Luo, W. B., Zhu, X. G., Gu, R. L., Li, H., & Luo, Y. J. (2014). Three stages of emotional word processing: An ERP study with rapid serial visual presentation. *Social Cognitive and Affective Neuroscience, 9*(12), 1897-1903.

Zhang, D. D., Luo, W. B., & Luo, Y. J. (2013). Single-trial ERP analysis reveals facial expression category in a three-stage scheme. *Brain Research, 15*(12), 78-88.

Zhang, D., Lin, Y., Liu, Y., Luo, Y., & Jiang, D. (2019). Memory encoding, retention and retrieval of disgusting and fearful faces. *Acta Psychologica Sinica, 51*(1), 36.

Note: Figure translations are in progress. See original paper for figures.