Does Working Memory Selectively Modulate Subjective Perception?

Wonyi Che[1,*] Tomoya Nakamura[1,2,3*] & Hakwan Lau[1,4,5]

[1]Center for Neuroscience Imaging Research, Institute for Basic Science, Suwon, Republic of Korea
[2]Center for Brain Science, RIKEN, Wako, Japan
[3]Japan Society for the Promotion of Science, Tokyo, Japan
[4]Department of Biomedical Engineering, Sungkyunkwan University, Suwon, Republic of Korea
[5]Department of Intelligent Precision Healthcare Convergence, Sungkyunkwan University, Suwon, Republic of Korea

*Correspondence: wychae@g.skku.edu; tomoya.nakamura@riken.jp

Abstract

This study aimed to investigate how working memory interacts with perception at the behavioral level. We examined whether working memory modulates subjective perception when the general capacity for perceptual processing is kept constant. This possibility is in part motivated by the neuropsychological phenomenon of blindsight, in which subjective perception and general perceptual capacity are dissociated. In a previous pilot study ($n = 8$), we manipulated the congruency between working memory and perceptual content. We observed a trend suggesting that holding categorically incongruent content in working memory selectively impaired sensitivity (d_a) in a Yes/No (Y/N) *detection* task, even when perceptual processing capacity was matched using a two-interval forced choice (2-IFC) task. In the current study ($n = 16$), we tested the effect of congruency on the Y/N *discrimination* task with an increased sample size. To match perceptual processing capacity across conditions, we titrated stimulus contrast in a two-interval forced choice (2-IFC) task again. Contrary to the previous pilot study, we did not find evidence that working memory selectively modulates subjective perception, as measured by sensitivity (d_a) in the Y/N task, when processing capacity in the 2-IFC task was matched. Given this discrepancy, we pre-register a follow-up data collection on the original pilot design (i.e., Y/N *detection*), including the analysis plan and stopping rule.

Introduction

Working memory engages overlapping neural mechanisms with perception, particularly within early visual areas such as V1 and V2 (Albers et al., 2013; Christophel et al., 2017). Neuroimaging studies (e.g., Gayet et al., 2017) have shown that maintaining the working memory contents enhances neural responses in the visual cortex when subsequent sensory input matches the memorized representation (a congruent stimulus). Given the shared neural mechanism, we may expect that working memory would likewise influence perception at the behavioral level.

Previous studies (Kang et al., 2011; Teng & Kravitz, 2019) have indeed shown that visual working memory not only actively interacts with early perceptual processing but can also bias subsequent perception. For example, Kang et al. (2011) demonstrated that participants who maintained a certain dot motion direction in working memory perceived subsequent motion stimuli as shifted away from the remembered direction—known as a repulsion effect. Similarly, Teng and Kravitz (2019) suggested that similarity between the content of the working memory and the perceptual task (i.e., color and orientation) systematically increased the degree of perceptual bias. Taken together, working memory can modulate perceptual processing.

However, it is unclear whether working memory mainly influences the general capacity for distinguishing between different stimuli, or whether it selectively influences the subjective experience of seeing. To address this question, we conducted a psychophysical experiment comparing perceptual performances in a two-interval forced-choice (2-IFC) and a Yes/No (Y/N) task, while the congruent or incongruent content was held in the working memory. In the 2-IFC task, one of two intervals contained a face stimulus, and the other contained a scrambled image matched in contrast, luminance, and size. In the Y/N task, either a face or a scrambled face was presented.

Both perceptual tasks occurred during the delay of the working memory task, and the trials were equally distributed across three congruency conditions (Stimulus Congruent: SC, Category Congruent: CC, Category Incongruent: CI; see Figure 1 for further details). Congruency was defined by whether the perceptual stimuli matched the content of the working memory task. In the congruent conditions, participants memorized faces, whereas in the category incongruent condition, they memorized scenes, while the perceptual tasks involved faces. We included both stimulus- and category-congruent conditions to establish a gradient of similarity between the working memory content and the perceptual input, allowing us to distinguish image-level (SC) from broader category-level (CC) influences on perceptual decisions.

The 2-IFC task is a standard measure of processing capacity, which we can consider as a relatively objective measure. It depends on one's ability to discriminate between a pair of external stimuli that is presented on every trial. Therefore, the

judgment is less likely to be influenced by subjective biases (but see Yeshurun et al., 2008). In contrast, the Y/N detection task can be considered a relatively more subjective task, as it requires a certain internal template or reference to compare with the input. Importantly, Azzopardi and Cowey (1997) found that Y/N sensitivity relative to 2-IFC was impaired in blindsight, suggesting that Y/N sensitivity may selectively track subjective (i.e., conscious) perception, at least in ways more than 2-IFC sensitivity does.

It has also been hypothesized that subjective experience may require some putative central mechanisms to distinguish internally generated representations (i.e., working memory and imagery) from incoming external perceptual inputs (Lau, 2019; Dijkstra & Fleming, 2023). Based on this hypothesis, one may expect working memory to selectively modulate subjective perception, rather than the general capacity of perceptual processing.

In a previous pilot study (n = 8), we observed a trend suggesting that working memory may selectively impair subjective perception in the Y/N *detection* task under categorically incongruent conditions, despite matched perceptual capacity in the 2-IFC task. Although this effect was not statistically robust, it motivated the current study to more rigorously test this possibility, using the Y/N *discrimination* task (in which the stimuli more closely matched those in our 2-IFC task) and a larger sample.

Therefore, in this study, we examined the interaction between working memory and perception at a near-threshold level to compare detection sensitivity measures between Y/N (discrimination) and 2-IFC (discrimination) tasks. Specifically, we hypothesized that the content of working memory selectively modulates subjective perception, without changing general processing capacity. That is, we predicted that working memory-perception congruence would have a greater effect on detection sensitivity in the Y/N task than on 2-IFC sensitivity.

Methods

Pre-registration

All analyses and experimental procedures described below were conducted in accordance with a pre-registered plan (https://www.biorxiv.org/content/10.1101/2025.11.14.688399v2.full).

Participants

Sixteen students (11 females, 5 males, mean age = 23.6 years, *SD* = 2.6) from Sungkyunkwan University participated in the study. The final sample size was determined based on an a priori power analysis using preliminary data. All were naïve to the purpose of the study. We pre-screened individuals registered in the SONA system (Center for Neuroscience Imaging Research Participation System), using the following eligibility criteria: self-reported normal or corrected-to-normal visual acuity, no color vision deficiency, and no prior psychiatric diagnosis. All participants provided written informed consent under procedures approved by the Institutional Review Board of Sungkyunkwan University.

Apparatus

The experiment was programmed and run using PsychoPy version 2024.2.4 (Peirce et al., 2019). Visual stimuli were presented on an HP p1230 gamma-calibrated CRT monitor with a resolution of 1024 × 768 pixels at a refresh rate of 60 Hz (width: 39.37 cm (900 pixels), 8-bit color depth, standard dynamic range). A CRT display was chosen because of its high temporal precision and minimal display latency, characteristics that make it well-suited for psychophysical experiments. In a darkened room, participants viewed the stimuli on the monitor at a fixed distance of 54.1 cm, with their head stabilized on the head-and-chin rest.

Stimuli

Stimuli consisted of faces, scenes, masks, and scrambled faces, all presented to participants in grayscale, on a uniform gray background (21 cd/m²). The face images were selected from the Chicago Face Database (CFD; Ma et al., 2015; http://www.chicagofaces.org), and the scene images were from the *Scene Size x Clutter Database* (Park et al., 2015). We used a pixelwise random noise image (generated and saved in advance) as a mask, which was presented right after each 2-IFC/YN stimulus to avoid afterimages. We generated scrambled face stimuli in Python by applying the Fourier image processing phase, which scrambles the original face images into quadrants and rearranges them. Original face images (from the Chicago Face Database; Ma et al., 2015) were first converted to grayscale and cropped to remove background regions. The two-dimensional Fast Fourier

Transform (FFT) of each image was then computed, and the phase spectrum was divided into four quadrants. Within each of the first and second quadrants, the phase values were randomly permuted along both the horizontal and vertical axes. The third and fourth quadrants were obtained by reflecting the permuted second and first quadrants across the horizontal axis, respectively. The permuted phase matrices were recombined with the original amplitude spectrum and inverse Fourier-transformed (IFFT) to reconstruct the scrambled images. This procedure preserves the global amplitude (i.e., spatial frequency and luminance distribution) of the original faces while disrupting configural information. The scrambled faces, therefore, served as controls, ensuring that low-level visual properties were matched while facial identity was removed.

Visual stimuli for the working-memory task (sample and probe) were presented at four isoeccentric peripheral locations (eccentricity: 11.7° of visual angle from the center of the monitor). The face stimuli for the 2-IFC and Y/N tasks were presented at the monitor center. In each experimental block, the same stimuli were shown only once, except as required by the task condition (i.e., perceptual and one of the working memory samples had to be the same in the stimulus congruent condition).

During the main experiment, the root-mean-square (RMS) contrast of the face stimuli in the 2-IFC task was adaptively controlled using a 1-up-3-down staircase procedure (Levitt, 1971), targeting a threshold of approximately 79% correct performance (implemented using the StairHandler function in PsychoPy). The starting contrast value was 2.3%, and the values were bounded between 0.005% and 14.1% on a logarithmic scale to avoid reaching zero contrast. We ran the staircase separately for each condition (SC, CC, CI; see Figure 1), and each staircase consisted of 33 trials to ensure condition-specific threshold values as well as comparable performance levels across conditions. The contrast value from the last (33rd) trial was taken as the final converged threshold, which was then used as the fixed contrast in the subsequent Y/N task under the same congruency condition.

Procedure

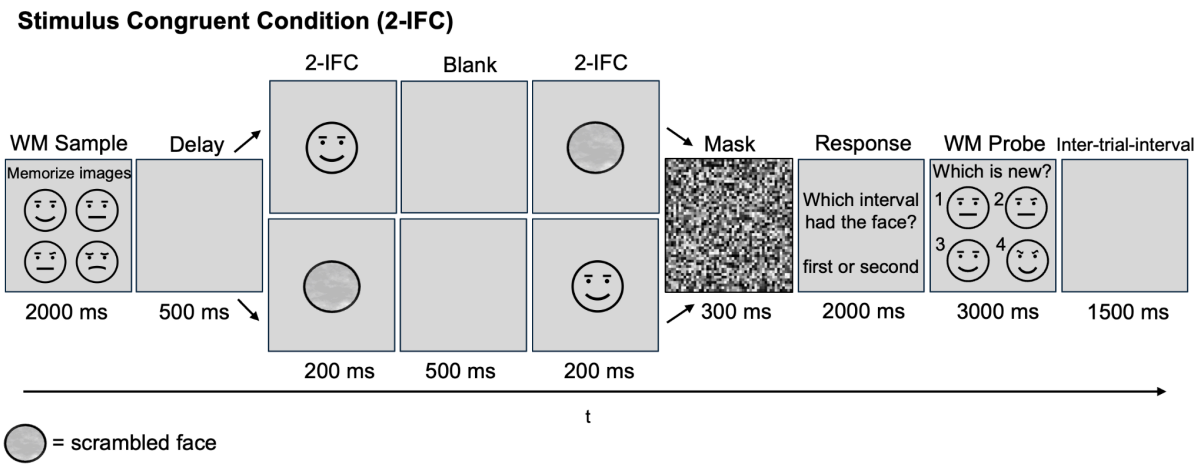
To examine how working memory content and its congruency systematically influence perceptual decisions, we used a 2 (perceptual tasks: 2-IFC, Y/N) × 3 (congruency: SC, CC, CI) within-subject factorial design. The experiment was conducted over two days, with three blocks (approximately 1.5h) completed per day. Each block consisted of two sub-blocks corresponding to the two perceptual tasks: the first sub-block was for the 2-IFC task, followed by the second sub-block for the Y/N task. Within each sub-block, 33 trials from each of three conditions (SC, CC, CI) were performed in random order. Participants were allowed to take an optional 5-minute break between blocks. In total, participants completed six such blocks, yielding 1,188 trials (6 blocks × 33 trials × 2 perceptual tasks × 3 congruency conditions).

On the first day, participants read the task instructions and completed 15 practice trials in the CC condition (3 trials for each: working memory task only, 2-IFC task only, Y/N task only, combined working memory-2-IFC, and combined working memory-Y/N) with trial-to-trial feedback. Feedback was shown as either “Correct” or “Incorrect”.

In the main experiment, participants were asked to perform either the 2-IFC or the Y/N task during the working memory delay period (Figure 1). At the beginning of each trial, as working memory sample stimuli, four different face images were presented for 2000 ms, followed by a 500-ms delay. In the 2-IFC sub-block, two stimuli (a face and a scrambled face) were randomly and sequentially shown for 200-ms each, with a 500-ms inter-stimulus interval. Following the 300-ms mask, participants had 2000-ms to indicate whether the face appeared in the first or second interval, using left or right arrow keys.

In the Y/N sub-block, either a face or a scrambled face was presented in a single interval, and participants judged whether a face appeared or not. Each face stimulus was followed by a 300-ms mask.

Regardless of the perceptual task, after the first response on the perceptual task, the working memory probe was presented: one of the four images was replaced from the initial sample sets, and the participants were asked to report which was the new image using the 1, 2, 3, or 4 keys within the time limit of 3000-ms. This probe display was followed by a 1500-ms inter-trial interval. A beep sound was played when the participant did not press any keys within the time limit. Nevertheless, the response was recorded up until the appearance of a working memory sample for the next trial, even after the beep sound. In the main experiment, no feedback was provided for either the working memory or the perception task.



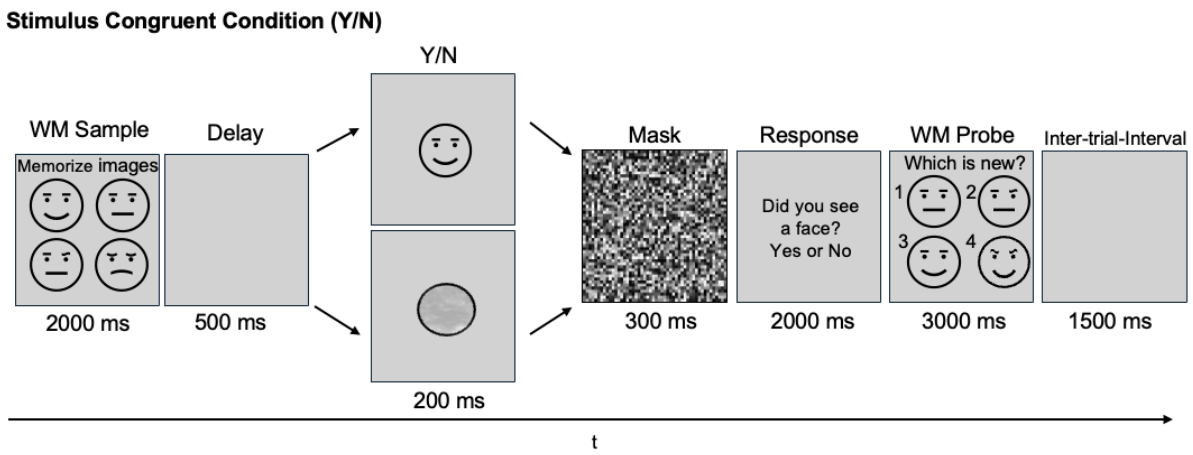


Figure 1. Experimental paradigm and task conditions *In the Stimulus Congruent (SC) condition, one of the working memory (WM) samples was physically identical to the face stimulus in the perceptual task (2-IFC and Y/N). In the Category Congruent (CC) condition, working memory stimuli belonged to the same category (faces) as the perceptual stimulus but were not physically identical. In the Category Incongruent (CI) condition, working memory stimuli belonged to a different category (scenes) from the perceptual face stimulus. Regardless of condition, a single trial consisted of a working memory sample display (encoding 4 images), a delay, a perceptual task (2-IFC or Y/N face detection), and finally a working memory probe display in which participants identified the novel stimulus. For copyright reasons, face stimuli were substituted in this manuscript with copyright-free emoji exemplars, only for illustrative purposes.*

Data Analysis

According to the criteria pre-registered in the previous version of preprint, participants who exhibited $d_a < 0$ in any condition or >20% missing responses were excluded and replaced.

Working memory task

We calculated the accuracy for the working memory task for each of the trials accompanied by 2-IFC and the Y/N task in the delay to examine whether the working memory performance varied depending on the type of perception task.

Perception task: stimulus detection

For the perception task, we calculated detection sensitivity, i.e., participants' ability to discriminate between faces (signals) and scrambled faces (noises), and criterion, according to signal detection theory (SDT; Macmillan & Creelman, 1991). All the following analyses were based on the unequal-variance SDT model because variances for the signal and noise distributions are usually different, especially for the Y/N task.

$$d_a = \sqrt{\frac{2}{1+s^2}}\,[\Phi^{-1}(HR) - s \cdot \Phi^{-1}(FAR)],$$

where Φ^{-1} denotes the z-transformation of hit rate (HR) and false alarm rate (FAR). For the Y/N task, HR was the proportion of “Yes” responses to the signal, and FAR was the proportion of “Yes” responses to the noise:

$$HR = P(\text{"}Yes\text{"}|signal)$$
$$FAR = P(\text{"}Yes\text{"}|noise)$$

To calculate the HR and FAR for the 2-IFC task, we treated the signal in the first interval as the “signal”:

$$HR = P(selecting\ the\ first\ interval \mid signal\ in\ the\ first\ interval)$$
$$FAR = P(selecting\ the\ first\ interval \mid signal\ in\ the\ second\ interval)$$

To avoid obtaining infinite z-values, 0 and 100% were adjusted by ± 0.5% prior to the z-transformation (Hautus, 1995). s represents the ratio of standard deviations of the signal and noise distribution ($s = \frac{\sigma s}{\sigma n}$). To get the estimate of s, we derived type-1 zROC (the z-transformed receiver operating characteristic function) using reaction time (RT) data (Miyoshi et al., 2025; see Supplementary Figure 1). Capitalizing on the empirical observation that quicker RT is associated with higher confidence, we divided trials into three RT bins (fast, middle, slow) for each response type. By shifting the decision criterion across these bins, we obtained five pairs of HR and FAR. A linear function was then fitted to the resulting zROC, and s was estimated as the inverse of the slope of the best-fitting line.

Perception task: metacognition

We quantified metacognitive sensitivity, i,e, the degree to which participants discriminated between their own correct and incorrect decisions, following the meta-SDT framework (Maniscalco & Lau, 2012). Instead of collecting participants' reported confidence levels, we inferred confidence from reaction times (RT), analogous to our approach in deriving type-1 zROC. We divided trials into three confidence bins (high, medium, and low), each for correct and incorrect responses. Metacognitive sensitivity ($meta\ d_a$) was estimated as d_a, which produces this type-2 contingency table with maximum likelihood (Maniscalco & Lau, 2012). The signal-to-noise ratio s was fixed at the value derived from type-1 zROC analysis (mentioned above) when fitting the meta-SDT model, to make the detection and metacognitive sensitivities comparable.

Lastly, metacognitive sensitivity was normalized by detection sensitivity to obtain metacognitive efficiency:

$$M_{ratio} = \frac{meta\ d_a}{d_a}$$

M_{ratio} reflects the extent to which metacognitive performance approaches the ideal level ($M_{ratio} = 1$) predicted from the individual's perceptual sensitivity. We presented it in the Supplementary Figure 1.

Results

For the working memory task, accuracy was well above chance (25%). Still, it did not reach ceiling levels regardless of the congruence condition or the perceptual task (2-IFC or Y/N) performed in the delay (Figure 2a), confirming that the participants understood and followed the task instruction reasonably well. In contrast, memory load was high enough to see the potential influence on the perceptual task. Specifically, the mean working memory accuracies accompanied by the 2-IFC task were 63.15%, 63.33%, and 64.80% for the SC, CC, and CI conditions, respectively. Working memory accuracies accompanied by the YN task were 67.62%, 67.16%, and 68.65%, respectively. Both showed a similar numerical trend, with accuracy being the highest in the category incongruent condition. However, this trend was not statistically significant ($F(2,30) = 0.60$, $p = .56$).

For the perception task, the accuracies of the 2-IFC and Y/N tasks were separately shown in Figure 2b. We calibrated the stimulus contrast in the 2-IFC task targeting ~79% accuracy, and mean accuracy was actually matched across congruence conditions as we expected (SC: 78.21%, CC: 76.02%, and CI: 76.94%). The Y/N accuracy (SC: 66.16%, CC: 63.87%, and CI: 66.00%) was above chance but lower than the 2-IFC accuracy ($F(1,15) = 95.50$, $p < .001$). However, neither the main effect of congruency ($F(2,30) = 1.74$, $p = .19$) nor the interaction ($F(2,30) = 0.21$, $p = .82$) was significant.

Figure 2c indicates the converged contrast threshold obtained from the staircasing during the 2-IFC task and subsequently used for the Y/N task. The average contrast thresholds were comparable across the conditions (SC: 3.9%, CC: 3.7%, and CI: 3.7%). A repeated-measures ANOVA revealed that the staircased contrast values did not significantly differ across conditions ($F(2,30) = 0.674$, $p = .516$, $\eta^2 G = .012$) and Bonferroni-corrected post-hoc pairwise comparisons showed no significant differences between any pairs (SC-CC: $p = .636$, CC-CI: $p = 1.00$, SC-CI: $p = 1.00$). Additionally, the 2-IFC sensitivity (d_a) did not differ across the conditions, $F(2,30) = 1.70$, $p = .200$, $\eta^2 G = .049$ (Figure 3a). These results indicate that stimulus intensity and processing capacity were matched across conditions.

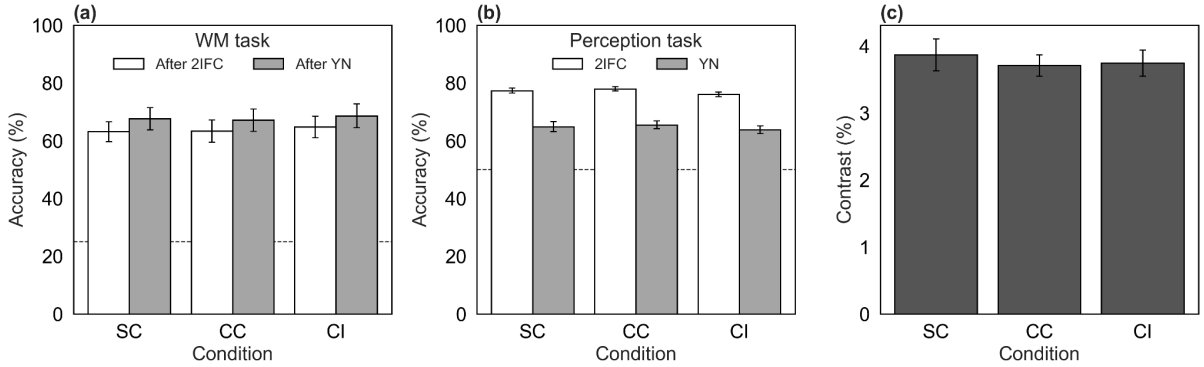


Figure 2. Group-level results of accuracy and contrast threshold *(a) Accuracies (%) for the working memory task, (b) Accuracies for the 2-IFC and Y/N task. (c) Average contrast thresholds derived from staircases in 2-IFC tasks.* Gray dashed lines indicate chance performance (25% for WM task; 50% for 2IFC and Y/N tasks). Error bars indicate standard errors of the mean.

Figure 3a shows that sensitivity (d_a) for the 2-IFC and Y/N task. A 2 x 3 repeated measure ANOVA showed a significant main effect of task ($F(1,15) = 43.89$, $p < .001$, $\eta^2 G = .307$), with higher sensitivity in 2-IFC than Y/N. In contrast, neither the main effect of the condition ($F(2,30) = 1.356$, $p = .273$, $\eta^2 G = .029$) nor the interaction effect ($F(2,30) = 0.188$, $p = .829$, $\eta^2 G = .002$) was significant. Additionally, we performed a one-way repeated-measures ANOVA on the YN/2-IFC detection sensitivity (d_a) ratio across congruence conditions, finding that the main effect of the condition was not significant ($F(2,30) = 0.172$, $p = .842$, $\eta^2 G = .007$).

In a 2-IFC task, participants have twice as many opportunities to view the stimulus as in a Y/N task, which theoretically increases sensitivity by a factor of √2 (Macmillan & Creelman, 1991; Azzopardi & Cowey, 1997):

$$\frac{2\text{-}IFC\ d_a}{\sqrt{2}} \approx YN\ d_a$$

After dividing the 2-IFC d_a by √2, it was similar to Y/N sensitivity (SC: scaled 2-IFC d_a = 1.077, Y/N d_a = 1.100; CC: scaled 2-IFC d_a = 1.122, Y/N d_a = 1.097; CI: scaled 2-IFC d_a = 1.028, Y/N d_a = 0.941). A 2 × 3 repeated-measures ANOVA confirmed that neither main effects (task: $F(1, 15) = 0.183$, $p = .674$, $\eta^2 g = .001$; condition: $F(2, 30) = 1.198$, $p = .315$, $\eta^2 g = .025$) nor interaction ($F(2, 30) = 0.28$, $p = .759$, $\eta^2 g = .004$) was significant. Therefore, our results were consistent with the theoretical relationship between Y/N and 2-IFC sensitivities.

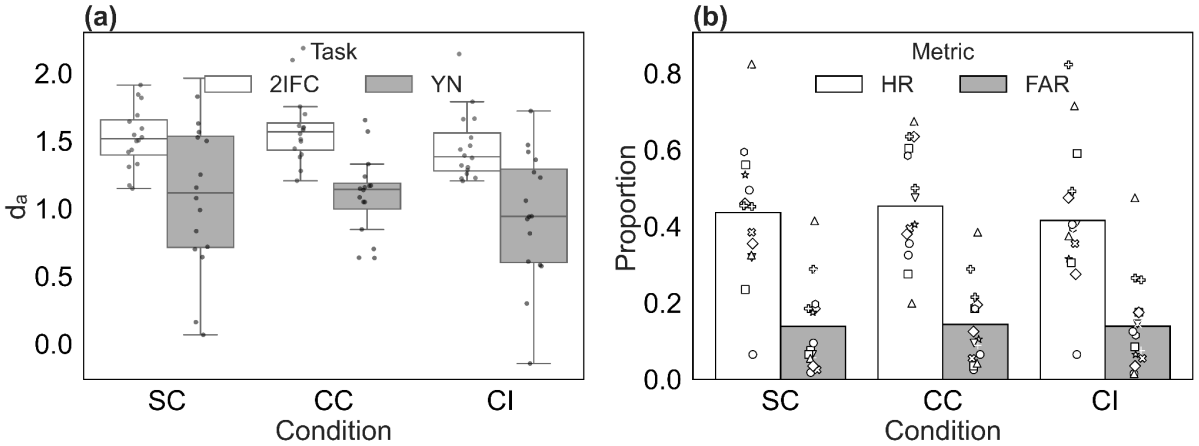


Figure 3. Group-level results of sensitivity in 2-IFC and Y/N tasks *(a) Detection sensitivity (d_a) in the 2-IFC and Y/N tasks. (b) Hit Rate (HR) and False Alarm Rates (FAR) in the YN task. Distinct markers indicate individual participants, and bar graphs indicate inter-individual averages.*

As shown in Figure 3b, when the stimulus in Y/N was incongruent with the working memory (i.e, CI), HR and FAR were the lowest (HR: 0.416, FAR: 0.130). This may reflect a more conservative criterion when the contents from working memory and perception contents are inconsistent or conflict. A 2 x 3 repeated-measures ANOVA confirmed that there was a significant main effect of the metric ($F(1,15) = 155.51$, $p < .001$), suggesting that the HR was significantly higher than the FAR. However, there was neither a main effect of condition ($F(2,30) = 0.38$, $p = .69$), nor the interaction between metric and conditions ($F(2,30) = 0.56$, $p = .58$).

Discussion

In this study, we investigated how working memory content influences perceptual decisions and whether its effect differs for objective processing capacity and subjective experience. Because the 2-IFC contrast values were titrated to maintain a constant accuracy within each working memory condition, processing capacity was matched. Applying these titrated contrasts to the Y/N task, therefore, allowed us to examine whether working memory modulates subjective detection independently of processing capacity. In this experiment, we used a scrambled face as the “noise” stimulus both in the Y/N and 2IFC tasks to better match stimulus and temporal predictability, thereby enabling a fairer comparison of detection sensitivity across perceptual tasks. We found that, when 2-IFC performance was titrated to be the same across the three different congruency conditions, working memory did not effectively modulate the Y/N sensitivity (d_a).

Despite statistical results, a slight numerical reduction in d_a was observed in the CI condition compared to SC and CC conditions (see Figure 3b). Importantly, this decrease, if any, occurred in both the 2-IFC and Y/N tasks, and the relative relationship between the two sensitivities remained stable across different congruency conditions. Thus, holding incongruent content in working memory perhaps reduced overall perceptual sensitivity but did not selectively affect subjective detection beyond its general effect on processing capacity.

Rather than revealing a dissociation between objective and subjective processing, the present findings may suggest a parallel reduction in perceptual sensitivity when participants held incongruent content in working memory (although this was not statistically validated with the current sample size). Furthermore, the theoretical relationship between sensitivity in Y/N and 2IFC tasks was not violated in this experiment. Ideally, the 2-IFC d_a is expected to be $\sqrt{2}$ times Y/N d_a**,** as long as all factors are matched except for the number of stimulus presentations (Macmillan & Creelman, 1991; but see Yeshurun et al., 2008). In our data, after scaling 2-IFC d_a by $\sqrt{2}$, we found that it remained indeed similar to the Y/N d_a in all of the 3 congruency conditions—differing by only 2.03% (SC), 2.22% (CC), 9.23% (CI), respectively. These patterns contrast with phenomena such as blindsight, in which objective discrimination can remain intact despite diminished subjective awareness (Azzopardi & Cowey, 1997). Instead, the current results may resemble classic visual suppression paradigms, such as pattern masking, where both discrimination sensitivity and subjective perception are deteriorated together (Knotts et al., 2018).

A closely related concept to working memory is mental imagery, which shares neural substrates with perception in the early visual cortex (Dijkstra et al., 2019). However, working memory reflects the maintenance of recently perceived information, whereas imagery may involve internally generated representations that are not necessarily tied to immediate sensory input. The present findings suggest that maintaining incongruent working memory content may modulate perceptual signal strength in a manner similar to other forms of sensory interference, rather than selectively disrupting subjective experience. Future work should investigate whether visual imagery interacts with perception in a similar way.

In the previous pilot study, subjective sensitivity in the Y/N task was selectively reduced under categorically incongruent conditions, suggesting that working memory modulates perception in a congruency-dependent manner. However, in the current study, sensitivity did not differ across congruency conditions, indicating that this selective effect was not replicated. The apparently inconsistent results may be related to differences in task design. One possible explanation is that the modified Y/N task design, in which scrambled faces were presented in target-absent trials, may have reduced sensitivity to congruency-dependent modulation. By introducing competing visual input in both target-present and target-absent trials, the task may have involved a more discrimination-like process rather than pure detection, thereby diminishing the influence of working memory content on subjective perception.

Given the divergence between the previous pilot and current results, we plan to return to the previous pilot design in which no stimulus is presented in target-absent trials in the Y/N paradigm. This pure detection version of the Y/N task may allow for a more direct assessment of whether working memory congruency selectively modulates subjective perception.

Here we pre-register additional data collection. To evaluate the robustness and replicability of the trend observed in the original sample of 8 participants in our earlier preprint, we will recruit an additional 8 new participants for that design. The results from the original 8 participants and the newly collected 8 participants will be reported to demonstrate the consistency of the effect across independent samples. Individual participants will be excluded if the HR is lower than the FAR, producing a negative sensitivity index (i.e., $da < 0$) at least in one condition, or if the data set contains no responses in more than 20% of the trials. Data collection will continue until we have 16 participants who do not meet these exclusion criteria. To test our main hypothesis, we will perform one-way repeated-measures ANOVA on the YN/2-IFC sensitivity ratio across congruence conditions in the 16 participants. If the main effect is significant, we will continue to perform two-tailed paired t-tests to examine pairwise differences (SC-CC, CC-CI, and SC-CI); p-values will be corrected by the Bonferroni method in these post-hoc tests.

In our earlier preprint, we proposed a follow-up experiment aimed at disentangling potential task-difficulty confounds from content-specific working-memory effects. Specifically, we planned to calibrate Y/N contrast levels (targeting ~79% accuracy) to match overall performance with the 2-IFC task, and to introduce a no–working memory baseline condition. This design was intended to test whether any apparent selective modulation of the Y/N task would persist when task difficulty was equated. However, the current result did not reveal a selective impairment of the Y/N task. Given the absence of evidence for a task-specific WM congruence effect, the original motivation for the planned difficulty-calibration experiment was substantially reduced. Thus, we decided not to proceed with that planned follow-up experiment.

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & De Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*(15), 1427–1431.

Azzopardi, P., & Cowey, A. (1997). Is blindsight like normal, near-threshold vision?. *Proceedings of the National Academy of Sciences*, *94*(25), 14190–14194.

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences*, *21*(2), 111–124.

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual Areas. *The Journal of Neuroscience*, 37(5), 1367–1373. https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Dijkstra, N., & Fleming, S. M. (2023).Subjective signal strength distinguishes reality from imagination. *Nature Communications,* 14, 1627 https://doi.org/10.1038/s41467-023-37322-1

Dijkstra, N., Hinne, M., Bosch, S. E., & van Gerven, M. A. J. (2019). Individual differences in the influence of mental imagery on conscious perception. *bioRxiv*, 607770. https://doi.org/10.1101/607770

Gayet, S., Guggenmos, M., Christophel, T. B., Haynes, J. D., Paffen, C. L. E., Vander Stigchel, S., & Sterzer, P. (2017). Visual Working Memory Enhances the Neural Response to Matching Visual Input. *The Journal of Neuroscience*, *37*(28), 6638–6647. https://doi.org/10.1523/JNEUROSCI.3418-16.2017

Hautus, M.J. (1995). Corrections for extreme proportions and their biasing effects on estimated values of *d′*. *Behavior Research Methods, Instruments, & Computers* 27, 46–51 https://doi.org/10.3758/BF03203619

Kang, M. S., Hong, S. W., Blake, R., & Woodman, G. F. (2011). Visual working memory contaminates perception. *Psychonomic Bulletin & Review*, *18*(5), 860–869. https://doi.org/10.3758/s13423-011-0126-5

Knotts, J. D., Lau, H., & Peters, M. A. K. (2018). Continuous flash suppression and monocular pattern masking impact subjective awareness similarly. *Attention,*

Perception & Psychophysics, *80*(8), 1974–1987. https://doi.org/10.3758/s13414-018-1578-8

Lakens, D., & Caldwell, AR. (2021). Simulation-Based Power Analysis for Factorial Analysis of Variance Designs. *Advances in Methods and Practices in Psychological Science*. 4(1). doi:10.1177/2515245920951503

Lau H. (2019) Consciousness, Metacognition, & Perceptual Reality Monitoring. *PsyArXiv*. https://doi.org/10.31234/osf.io/ckbyf

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49*(2, Pt. 2), 467-477.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, 47, 1122–1135. https://doi.org/10.3758/s13428-014-0532-5

Macmillan, N. A., & Creelman, C. D. (1991). Detection Theory: A user's guide. Cambridge: Cambridge University Press.

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

Miyoshi, K., Rahnev, D., & Lau, H. (2025). Response Time as Decision Confidence: Insights from Type-2 ROC Analysis. https://doi.org/10.31234/osf.io/6gyjf_v1

Park, S., Konkel, T., & Oliva, A. (2015). Parametric coding of the size and clutter of natural scenes in the human brain. *Cerebral Cortex*, 25(7), 1792–1805. https://doi.org/10.1093/cercor/bht418

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. (2019). PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*. 10.3758/s13428-018-01193-y

Teng, C., & Kravitz, D.J. (2019). Visual working memory directly alters perception. *Nature Human Behaviour* 3, 827–836. https://doi.org/10.1038/s41562-019-0640-4

Yeshurun, Y., Carrasco, M., & Maloney, L. T. (2008). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. *Vision Research*, *48*(17), 1837–1851. https://doi.org/10.1016/j.visres.2008.05.008