Neuropsychologia 192 (2024) 108745

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia

The role of visual working memory in capacity-limited cross-modal ensemble coding

Greer Gillies [a,*], Keisuke Fukuda [b], Jonathan S. Cant [b]

[a] University of Toronto, Canada
[b] University of Toronto, Mississauga Campus, Canada

ARTICLE INFO

Keywords:
ensemble coding
Vision
Taste
Cross-modal
Visual working memory

ABSTRACT

Ensemble coding refers to the brain's ability to rapidly extract summary statistics, such as average size and average cost, from a large set of visual stimuli. Although ensemble coding is thought to circumvent a capacity limit of visual working memory, we recently observed a VWM-like capacity limit in an ensemble task where observers extracted the average sweetness of groups of food pictures (i.e., they could only integrate information from four out of six available items), thus suggesting the involvement of VWM in this novel form of cross-modal ensemble coding. Therefore, across two experiments we investigated if this cross-modal ensemble capacity limit could be explained by individual differences in VWM processing. To test this, observers performed both an ensemble task and a VWM task, and we determined 1) how much information they integrated into their ensemble percepts, and 2) how much information they remembered from those displays. Interestingly, we found that individual differences in VWM capacity did *not* explain differences in performance on the ensemble coding task (i.e., high-capacity individuals did not have significantly higher "ensemble abilities" than low-capacity individuals). While our data cannot definitively state whether or not VWM is *necessary* to perform the ensemble task, we conclude that it is certainly not *sufficient* to support this cognitive process. We speculate that the capacity limit may be explained by 1) a bottleneck at the perceptual stage (i.e., a failure to process multiple visual features across multiple items, as there are no singular features that convey taste), or 2) the interaction of multiple cognitive systems (e.g., VWM, gustatory working memory, long term memory). Our results highlight the importance of examining ensemble perception across multiple sensory and cognitive domains to provide a clearer picture of the mechanisms underlying everyday behavior.

Our perception of the visual world is richly detailed, but this subjective experience is at odds with a body of research that suggests that what we are capable of perceiving is limited by a number of cognitive systems such as attention and visual working memory (VWM) (Cohen et al., 2016; Luck and Vogel, 1997). However, while our visual environments are dense with information, this information is not completely random, with scenes containing groups of similar objects and features (Whitney and Yamanashi Yamanashi Leib, 2018). Our visual system is highly sensitive to these regularities and to process them, the brain can utilize ensemble coding, which is the ability to represent large amounts of information from groups of similar items as a single summary statistic (e.g., the average expression of a crowd of faces). Ensemble coding helps us circumvent the encoding bottleneck of VWM (Khayat and Hochstein, 2019), which can only hold about 3 or 4 items at a time (Luck and Vogel, 1997).

Despite the known relationship between VWM and ensemble coding, and to borrow a term used by Melvyn Goodale when referring to the relationship between the dorsal and ventral streams of visual processing (Goodale, 2011), ensemble coding and VWM are not 'hermetically sealed' from one another, and instead these two seemingly distinct cognitive systems routinely interact. In this study we investigated the underlying nature of a possible interaction between VWM and ensemble coding.

1. Interactions between VWM and ensemble coding

Much of the research on interactions between ensemble coding and VWM has focussed on how the former influences the latter. For example, Brady and Alvarez (2011) (see also Papenmeier and Timm, 2021) found that an individual's VWM representations could be biased by ensemble

* Corresponding author. Department of Psychology, Toronto, Ontario, Canada.
E-mail address: greer.gillies@mail.utoronto.ca (G. Gillies).

https://doi.org/10.1016/j.neuropsychologia.2023.108745
Received 28 June 2023; Received in revised form 1 December 2023; Accepted 4 December 2023
Available online 13 December 2023

statistics. Observers viewed sets of differently sized colored circles (red and blue) and were told to remember the size of all circles of one color. When participants reported the color of one of the probed circles, their reports of the single item were biased in the direction of the average size of all circles of that color.

Such biases in VWM representations towards the mean of groups of items have been found across a range of different stimuli such as faces (Griffiths et al., 2018; Corbin and Crawford, 2018) and oriented triangles (Utochkin and Brady, 2020). Together, these results reveal that summary statistics from ensembles of multiple objects can bias memory representations for single items.

More recently, research has focussed on the inverse relationship, that is, the influence that VWM can have on ensemble coding. Williams et al. (2021) found that holding a single item in VWM (e.g., a colored shape) biased performance on a subsequent ensemble task. Observers performed reported the average orientation of a group of lines (presented in two colors) in tandem with a VWM task (i.e., remember a colored shape). Reports for average line orientation were biased in the direction of the lines that matched the color of the shape held in VWM (though this color was irrelevant to the ensemble task). This demonstrates that the contents of VWM can bias ensemble representations.

Another study by Brand et al. (2012) found that having a VWM template can aide performance on an ensemble task when the to-be-averaged items match a feature of that template. Participants were shown differently sized circles in two colors and were asked to report the average size of the circles in one of the colors. On some trials, participants were cued as to what color would be tested, and participants had improved accuracy in the cued compared to the un-cued condition. In these cases, VWM representations can be detrimental to or aide performance on an ensemble task, depending on what features are relevant in the ensemble display. Together, these studies clearly demonstrate the bi-directional influence that VWM representations and ensemble coding have on each other, thus revealing the routine interactions that take place between these different cognitive systems.

2. Capacity limits in ensemble coding

Although there are bi-directional interactions between VWM and ensemble coding, it is still unclear if there are cases when VWM may have a more direct involvement in ensemble coding (e.g., the degree to which VWM contents can be used to generate, rather than simply bias, ensemble percepts). Although ensemble coding is thought to circumvent capacity limits to VWM, capacity limits have been observed in ensemble coding tasks (Maule and Franklin, 2016; Ji et al., 2018). For example, Maule and Franklin (2016) found that observers were subsampling (i.e., using only some of the available information) from arrays of colored circles when engaging in an average color task. Specifically, they found that a model that randomly subsampled two out of 16 available circles produced results equivalent to that of the participants. Similarly, Ji et al. (2018) found a capacity limit for the number of faces that could be integrated in an average facial expression task, as task performance reliably decreased as the set size of the face stimuli increased.

Importantly, studies that encountered capacity limitations or found that participants were engaging in subsampling strategies did not investigate the possible involvement of VWM in such processes (e.g., participants might have been subsampling from the array and holding those items within VWM to then generate an ensemble percept). As the capacity limitations found in the aforementioned ensemble studies are similar to what has been found with VWM (i.e., less than 4 items), investigating the possible role of VWM in explaining these ensemble capacity limitations is critical.

Recently, we encountered a capacity limit in a cross-modal ensemble coding task where observers were asked to extract the average taste (i.e., sweetness) from visually presented food items (Gillies et al., 2023). Notably, visual features alone do not fully convey the taste of different foods, and thus to make this judgement, observers must use visual information to cue knowledge of taste stored in long term memory (LTM). We found evidence that observers had a limited ability to perceive the average sweetness of visually presented foods. Of interest, under simultaneous viewing conditions (all stimuli presented at once), observers were limited in the number of items they could incorporate into their cross-modal ensemble percepts. Specifically, when observers were shown six food pictures, they could only use information from four items (i.e., they may have been subsampling due to a capacity limitation). This four-item capacity limit persisted even with increased viewing time of the ensemble displays (1s and 1.5s).

This capacity limit is unlike other forms of “abstract” ensemble coding studied previously (i.e., average object animacy, average economic value of items; Yamanashi-Leib et al., 2016, 2020), even though those studies also used complex stimuli (i.e., multifeatured objects) and in a paradigm that inspired the experimental design we used in Gillies et al. (2023). For both object animacy and economic value, observers were able to integrate information from all six available items to extract summary statistics under brief viewing durations, showing no evidence of subsampling from the displays.

The possible involvement of VWM in this ensemble coding task is worth investigating, given the capacity limit. Indeed, the four-item capacity limit in Gillies et al. (2023) may be reflective of a VWM capacity limit, which is also limited to about four items (e.g., Luck and Vogel, 1997). The viewing times used in Gillies et al. (2023) were certainly long enough time to make several saccades (1 and 1.5s), and thus participants could have serially fixated on a subsample of the items in the displays, and then used the information held in VWM to aid in the formation of the ensemble percept. Here we conduct two experiments to investigate the role of VWM in explaining the capacity limitation observed in the ensemble coding of average sweetness.

3. The current study

If participants are using VWM to aide their performance on the ensemble sweetness task, then they should have explicit memory for the items within the ensemble display. To investigate this possibility, we had observers perform both an ensemble sweetness task and a VWM task to ascertain 1) how many items they integrated into their ensemble percepts, and 2) how much information they remembered from the displays. If VWM fully supports performance on the ensemble task, then individuals with higher VWM capacities should therefore show improved performance on the ensemble task compared to lower capacity individuals. To preview our results, we found that individual differences in VWM performance could *not* predict differences in performance on the ensemble task.

4. Experiment 1

The purpose of Experiment 1 was to investigate if VWM explains performance on the ensemble sweetness task from Gillies et al. (2023).

For the ensemble task, participants were shown visual arrays containing multiple food pictures and were asked to rate their average sweetness. The food pictures used in this experiment and Experiment 2 were previously validated (see supplementary materials for Gillies et al., 2023). That is, an independent group of raters viewed the 150 food pictures one-at-at-time and rated them on their perceived sweetness on a scale from 0 (not sweet at all) to 10 (extremely sweet). Inter-rater reliability indexed by an intraclass correlation coefficient was high, ICC = 0.98 (Cicchetti, 1994), demonstrating that participants reliably rated perceived sweetness, and, importantly, that there was excellent agreement across these raters in their sweetness percepts of individual food items (Gillies et al., 2023).

To determine how much information observers were integrating from the food ensembles, we used a subset manipulation (Yamanashi-Leib et al., 2016, 2020; Gillies et al., 2023). On some trials, observers were only shown a subset of the whole ensemble (one, two, or four items from the six-item ensemble). Importantly, their responses on these subset conditions were only ever compared to the predicted sweetness of the full ensemble (derived from the individual food ratings conducted by the independent group of observers in Gillies et al., 2023). Specifically, for each participant and set size, we calculated the correlation between the participants' reported sweetness ratings and the predicted sweetness ratings for the full 6-item ensemble. As the subsets are not representative of the predicted sweetness of the full ensemble, this analysis simulates a subsampling strategy. If participants are only using some of the available information to make their ensemble judgements, the correlation between participants' actual and predicted sweetness ratings would plateau at smaller subset sizes (see Fig. 1). If, however, participants are able to integrate information from multiple items, the correlation would increase with increasing subset size. We predict that we will replicate the 4-item capacity limit found in Gillies et al. (2023).

For the VWM task, participants were again shown the same

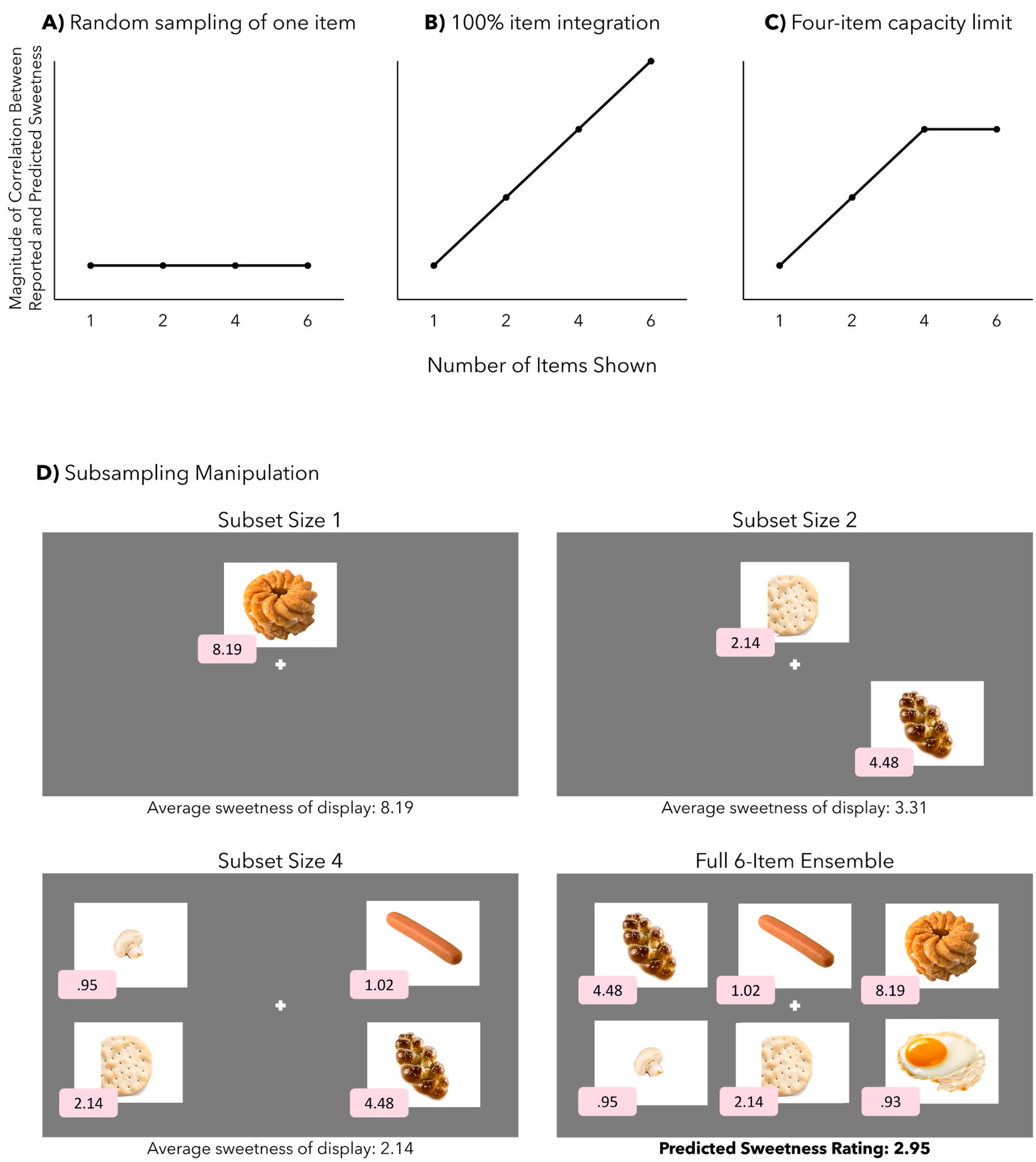

Fig. 1. Subset Manipulation

A) The pattern of results that would occur if participants sampled only one item from an array. This pattern would show that observers are unable to integrate information from multiple items when making their ensemble judgements.

B) The pattern of results that would occur if participants successfully integrated all six items in the array. Here, the correlation would increase at larger set sizes as more information becomes available to participants. This pattern would show that observers could use all the information available to them when making their ensemble judgements.

C) The predicted pattern of results where the magnitude of the correlation increases until subset size 4.

D) An illustration of the subsampling manipulation used. Shown in the pink boxes are the sweetness scores of each individual food picture. As more information is made available, the average sweetness of the display gets closer to the predicted sweetness rating of the full six-item ensemble.

ensembles as in the ensemble task but were then presented with a two-alternative forced choice (2AFC) membership identity task (see Yamanashi Leib et al., 2016, 2020). After the ensemble display disappeared, two items were presented (one from the previous ensemble and one from a different ensemble), and participants indicated which one they remembered seeing previously. From this, we can calculate how VWM capacity (the number of items recognized) changes with set size (see 'Analysis 2AFC task' below).

If VWM performance explains performance on the ensemble task, we should see an increase in VWM capacity as subset size increases, and, like performance on the ensemble task, this increase would plateau at subset size four. Moreover, if VWM fully supports performance on the ensemble task, then we should see a relationship between VWM capacity and ensemble coding abilities. Specifically, individuals with higher VWM capacities should be able to incorporate more information into their ensemble percepts. If we see evidence for both hypotheses, then we can conclude that VWM explains performance on the ensemble sweetness task. If, however, we observe a dissimilar pattern of results across ensemble integration and VWM capacity and/or no relationship between ensemble performance and VWM capacity, then we can conclude that VWM does not explain how observers perform on the ensemble sweetness task.

5. Methods

5.1. Participants

Participants were recruited via Prolific (2021), an online data-collection platform. Participants were pre-screened to ensure they met the following criteria: they currently reside in the US or Canada, were fluent in English, were between the age of 18–40 years old, had no head injuries, no ongoing mental health issues or illness, no cognitive impairments or dementia, and had normal or corrected-to-normal vision. They were paid $15.78 CAD per hour. As the experiment took 15–20 min to complete, most participants made approximately $4.00 CAD. Each participant provided electronic consent to the protocol approved by the Research Ethics Boards of the University of Toronto prior to participation.

A total of 88 participants were recruited, and six were excluded (see 'Participant Exclusion Criteria' below for details), leaving a final sample size of 82. An a-priori power analysis was conducted using G*Power3 (Faul et al., 2007) to determine how many participants would be needed for a correlation analysis with a medium effect size ($r = 0.30$). The results showed that a sample size of 82 would be required to obtain a power of .80.

The mean age of the final sample was 29.01 years, with 46 females, 35 males, and one who declined to answer. Seventy-six participants were right-handed, and six were left-handed. All had normal or corrected-to-normal vision, with 30 wearing glasses, 13 wearing contacts, and the rest requiring neither.

5.2. Participant Exclusion Criteria

To ensure that participants were not randomly clicking on the average sweetness rating scale (see 'Stimuli' below), their responses on the subset size one condition were correlated with the predicted sweetness values of the single items. Given that previous results showed that participants were in high agreement with one another as to how sweet the individual food items were (Gillies et al., 2023), it is reasonable to predict that the current participant ratings would be highly correlated with the individual sweetness scores. Indeed, the average correlation between the two was $r = 0.83$ (without exclusions). Participants who had a correlation below an r of 0.70 were excluded from further analysis. Using this criteria, six participants were removed from further analysis, leaving a final sample size of 82.

5.3. Apparatus

All data were collected online. Participants were directed from Prolific to Qualtrics (2020), where they read and digitally signed the consent form and answered demographics questions. After submitting the Qualtrics survey, they were redirected to Pavlovia (Peirce et al., 2019) which was used to run the experiment. The experiment was coded using Psychopy3 (Peirce et al., 2019). Both Mac and Windows machines were permitted (desktop or laptop). As this experiment was conducted online, the distance between the observer and the screen could not be reliably controlled, but participants were always asked to sit about an arms distance from their computer screen. In addition, participants were instructed to perform the experiment in a distraction-free environment, with their computer plugged in and charging and the screen brightness set to maximum. They were asked to maintain fixation on a central cross throughout the experiment.

5.4. Stimuli

Stimuli were taken from the Food-Pics database (Blechert et al., 2019) and the FoodCast research image database (FRIDa) (Foroni et al., 2013). The image database we put together contained 127 pictures from the Food-Pics database (Blechert et al., 2019), 34 of which were edited in photoshop (Adobe Photoshop, 2004) to ensure that only a single food item was present in each picture. We used an additional 23 pictures from the FRIDa database (Foroni et al., 2013), four of which were edited in Photoshop. The final 150 food stimuli were taken from a larger pool of food pictures (see Gillies et al., 2023, Supplementary Materials). Over a series of pilot studies, participants rated the perceived sweetness of food pictures, and we selected 150 images (from an initial 547 pictures) across a broad range of sweetness ratings with low standard deviations of response. Importantly, an intraclass correlation coefficient (ICC) score of 0.98 (Cicchetti, 1994) showed that observers were in high agreement with one another.

Ensemble Creation. From the 150 images, we randomly drew 6 images without replacement, yielding 25 sets of images with 6 images per set. Each set was assigned a predicted sweetness rating, calculated by averaging the individual ratings of the six items within the set (see Gillies et al., 2023). No one item in the set was within 0.50 units of the predicted average sweetness rating. The ensembles' predicted sweetness ratings were normally distributed around a mean of 4.93. For the subset conditions, one, two, or four items were randomly drawn from the full set (with replacement).

The ensemble arrays were presented in a 3×2 grid in the middle of the screen on a grey background (see Fig. 2). The location of each item was randomly determined within the grid. Each stimulus was 0.30×0.225 times the screen's height. Each item in the grid was separated both horizontally and vertically by 0.05 times the screen's height. A white fixation cross (0.04×0.04 times the screen's height) was presented in the middle of the screen.

For the ensemble task, a clickable rating scale was used to obtain participant ratings after the presentation of the ensemble. The scale ranged from 0 to 10, and the numbers were presented on the scale below 21 corresponding tick marks (each half value was represented by a tick mark). Participants made their ratings by clicking directly on the scale. Scale granularity was set to 0.25, to allow participants to use whole values, half values, and quarter values. The instructions above the scale read "On average, how sweet were those foods? Click on the rating scale to make your response. 0 = not sweet at all, 10 = extremely sweet." Participants were encouraged to use the full range of the scale.

For the VWM task, two items were shown on the left and right of the central fixation cross after the presentation of the ensemble (see Fig. 2). The images were separated by .10 times the screens height. Above the images in white font were instructions reading "Which food do you remember seeing?" and participants used their left and right arrow keys to select an item.

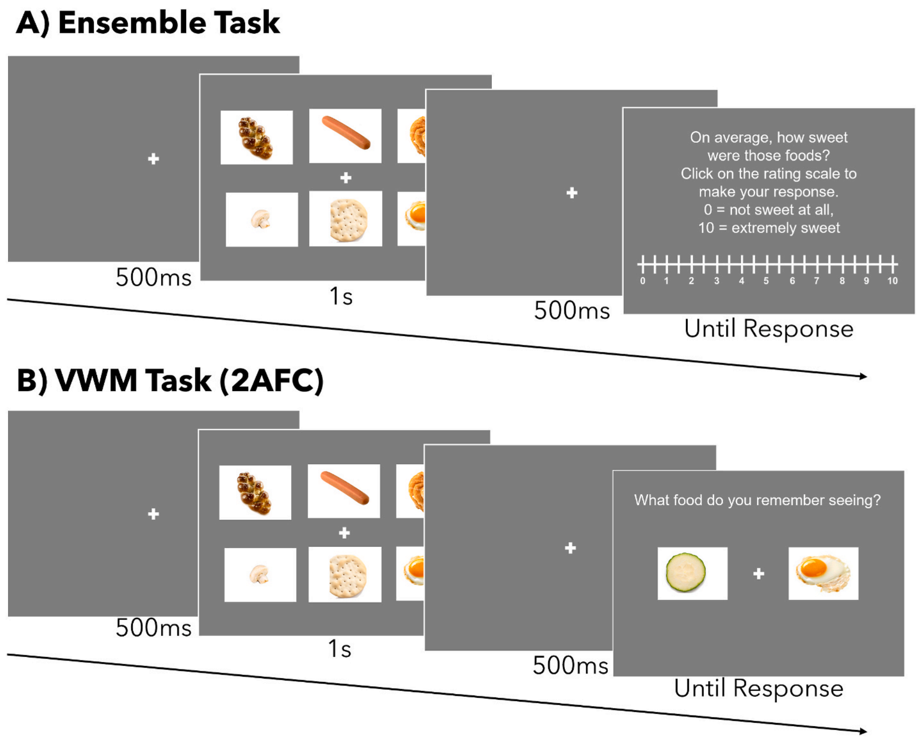

Fig. 2. Ensemble Task and VWM Task Sequences
A) The ensemble task used in both Experiments 1 and 2. This is an example of the full six-item ensemble condition. Participants viewed the images for 1s and then made an average sweetness rating using a rating scale from 0 (not sweet at all) to 10 (extremely sweet).
B) The VWM task (2AFC) used in Experiment 1. This is an example of the full six-item condition. Participants viewed the images for 1s, and were then shown two images, and indicated which image was from the previous set.

5.5. Procedure

Task order was counterbalanced between participants. Both the ensemble and VWM tasks began the same way. Participants were asked to maintain fixation on a central cross for 500ms and were then shown the full six-item ensemble or a subset of the full ensemble (one, two, or four items) for 1 s, followed by a 500ms delay in which only the fixation cross was present.

For the ensemble task, participants were instructed to make judgements about the average sweetness of groups of food items that could vary in set size and were asked to maintain fixation on the central cross. Following the delay, the rating scale appeared, and participants clicked directly on the rating scale to make their average sweetness judgements. The rating scale was present until a response was made (via mouse click). Participant responses were not meant to be speeded, and as such, participants were not instructed to respond as quickly as possible. Thus, we did not examine reaction time as a dependent measure in this study. Participants saw all 25 ensembles at each set size, for a total of 100 trials for the ensemble task.

For the VWM task, participants were instructed to remember the individual items from groups of food pictures that could vary in set size, while maintaining fixation on a central cross. Following the 500ms delay, participants were then shown two items on the left and right sides of the central cross, respectively. One of the items was from the previous ensemble display (the target), and the second was an item from one of the other ensembles (selected randomly) (the distractor). The positions of the target and distractor were randomly selected. Using the left and right arrow keys on their keyboards, participants indicated which food picture they remembered seeing. Both the target and distractor were present until the participant made a response. As in the ensemble task, responses were not speeded. Participants saw all 25 ensembles at each set size for a total of 100 trials for the VWM task.

6. Analysis

6.1. Ensemble task

To examine performance on the ensemble task, for each participant and set size condition, we calculated the correlation between the participants' reported sweetness ratings and the predicted sweetness ratings of the whole set (i.e., the 6-item ensemble). Next, to normalize the distribution of Pearson coefficients, Pearson correlations were converted to Fisher *Z* scores. This left us with four Fisher *Z* scores per participant, one for each set size. We then examined the relationship between the magnitude of the correlation (Fisher Z scores) and the number of items shown to individuals using a linear regression.

This linear regression was then followed up with three planned paired-sample *t* tests (i.e., comparing set size 1 to 2, 2 to 4, and 4 to 6). This analysis enabled us to identify if there was a significant increase in the magnitude of the correlation between all the subset conditions (i.e., if participants were incorporating more information as it was made available to them), or if the magnitude plateaus after a certain subset size (i.e., not all the information was incorporated into their ensemble judgements). To correct for multiple comparisons between the planned paired-sample *t*-tests, the Bonferroni corrected alpha value was set to 0.016 (i.e., for three comparisons). As we anticipated null results (i.e., no significant difference between set sizes 4 and 6), we also report Bayes factors for each comparison. Specifically, we report BF_{10} for significant results, and values greater than 3 can be interpreted as substantial evidence in favor of the alternative hypothesis (Wagenmakers et al., 2011; Jeffreys, 1961). For non-significant results, we report BF_{01}, and values greater than 3 can be interpreted as substantial evidence in favor of the null hypothesis (Wagenmakers et al., 2011; Jeffreys, 1961). See Tables 1 and 2 for a summary of the Bayesian results for Experiments 1 and 2, respectively.

Table 1
Summary of Results in Experiment 1 A summary table of the frequentist test results and Bayesian results for all the comparisons made for the ensemble and VWM tasks.

Comparison	Direction of Effect	Frequentist Test Result	Bayes' Factor	Strength of Evidence
Experiment 1 Ensemble Task Bayesian paired sample *t*-tests				
Set size 1 to 2	Increase	Significant	BF_{10} = 74.78	Very strong evidence for the alternative hypothesis
Set size 2 to 4	Increase	Significant	BF_{10} = 1672.74	Extreme evidence for the alternative hypothesis
Set size 4 to 6	No difference	*ns*	BF_{01} = 8.17	Substantial evidence for the null hypothesis
VWM Task Bayesian paired sample *t*-tests				
Comparison	Direction of effect	Frequentist Test Result	Bayes' Factor	Strength of evidence
Set size 1 to 2	Increase	Significant	BF_{10} = 619,989.91	Extreme evidence for the alternative hypothesis
Set size 2 to 4	Increase	Significant	$BF_{10} = 3.09 \times 10^{12}$	Extreme evidence for the alternative hypothesis
Set size 4 to 6	No difference	*ns*	BF_{01} = 1.41	Anecdotal evidence for the null hypothesis
VWM Task Split Bayesian independent samples *t*-tests				
Comparison (low vs. high-capacity participants)	Direction of effect	Frequentist Test Result	Bayes' Factor	Strength of evidence
VWM Ability	Increase	Significant	$BF_{10} = 6.83 \times 10^{15}$	Extreme evidence for the alternative hypothesis
Ensemble Ability	No difference	*ns*	BF_{01} = 4.13	Substantial evidence for the null hypothesis

6.2. VWM task

For each participant and set size condition (1, 2, 4, 6), we calculated the proportion of correct responses. Next, we determined each participant's VWM capacity for each set size condition using the following formula:

$$\text{VWM capacity} = (\text{current set size} \ * \ (\text{proportion correct} \ - \ .50) \ * \ 2) \ (\text{Yamanashi} - \text{Leib et al., 2016}).$$

As in the ensemble task, we ran a linear regression to examine the relationship between VWM capacity and set size, followed by the same three planned paired-sample *t*-tests (again corrected for multiple comparisons).

6.2.1. VWM split

To further examine the relationship between VWM capacity and ensemble coding, we median-split participants into high capacity and low-capacity groups based on their average VWM capacity at set sizes four and six. To establish the reliability of this measure, we then performed a split-half analysis. Responses were ordered by set size and were then randomly assigned an odd or even value. Therefore, one half of the split had 12 trials per condition, and the other half had 13.

Next, we defined measures of "VWM ability" (the average VWM capacity at set size 1 and 2 subtracted from the average VWM capacity at set size 4 and 6) and "ensemble ability" (the average Fisher Z at set size 1 and 2 subtracted from the average Fisher Z at set size 4 and 6).

Table 2
Summary of Results in Experiment 2 A summary table of the frequentist test results and Bayesian results for all the comparisons made for the ensemble and VWM tasks.

Comparison	Direction of effect	Frequentist Test Result	Bayes' Factor	Strength of evidence
Experiment 2 Ensemble Task Bayesian paired *t*-tests				
Set size 1 to 2	Increase	Significant	BF_{10} = 12.71	Strong evidence for the alternative hypothesis
Set size 2 to 4	Increase	Significant	BF_{10} = 79.20	Very strong evidence for the alternative hypothesis
Set size 4 to 6	No difference	*ns*	BF_{01} = 2.54	Anecdotal evidence for the null hypothesis
VWM Task Bayesian paired *t*-tests				
Comparison	Direction of effect	Frequentist Test Result	Bayes' Factor	Strength of evidence
Set size 1 to 2	Increase	Significant	BF_{10} = 262,909.79	Extreme evidence for the alternative hypothesis
Set size 2 to 4	Increase	Significant	BF_{10} = 84,14.28	Extreme evidence for the alternative hypothesis
Set size 4 to 6	No difference	*ns*	BF_{01} = 5.02	Substantial evidence for the null hypothesis
VWM Task (lenient grading) Bayesian paired *t*-tests				
Comparison	Direction of effect	Frequentist Test Result	Bayes' Factor	Strength of evidence
Set size 1 to 2	Increase	Significant	BF_{10} = 42,824.47	Extreme evidence for the alternative hypothesis
Set size 2 to 4	Increase	Significant	$BF_{10} = 5.22 \times 10^{6}$	Extreme evidence for the alternative hypothesis
Set size 4 to 6	Increase	Significant	BF_{10} = 64.25	Very strong evidence for the alternative hypothesis
VWM Task Split Bayesian independent samples *t*-tests				
Comparison (low vs. high-capacity participants)	Direction of effect	Frequentist Test Result	Bayes' Factor	Strength of evidence
VWM Ability	Increase	Significant	BF_{10} = 11,286.45	Extreme evidence for the alternative hypothesis
Ensemble Ability	No difference	*ns*	BF_{01} = 4.03	Substantial evidence for the null hypothesis
VWM Task Split (lenient grading) Bayesian independent samples *t*-tests				
Comparison (low vs. high-capacity participants)	Direction of effect	Frequentist Test Result	Bayes' Factor	Strength of evidence
VWM Ability	Increase	Significant	BF_{10} = 3095.64	Extreme evidence for the alternative hypothesis
Ensemble Ability	No difference	*ns*	BF_{01} = 2.83	Anecdotal evidence for the null hypothesis

Differences between high and low-capacity individuals on both measures were examined with two independent sample *t*-tests.

7. Results and discussion

7.1. Ensemble task

The average Fisher *Z* scores were fit by a linear regression, $r^2 = 0.14$, $p < .001$, indicating that participants were getting closer to the predicted sweetness rating as more information was made available to them (see Fig. 3A).

The planned paired *t*-tests revealed that there was a significant increase in Fisher *Z* scores between set size one and two [$t\,(81) = 3.79$, $p < .001$, Cohen's $d = 0.42$, 95% *CI* [0.19, 0.64], $BF_{10} = 74.78$], and between set size two and four [$t\,(81) = 4.71$, $p < .001$, Cohen's $d = 0.52$, 95% *CI* [0.29, 0.75], $BF_{10} = 1672.64$]. However, there was no significant difference in Fisher *Z* from set size four to six [$t\,(81) = 0.09$, $p = .93$, Cohen's $d = 0.01$, 95% *CI* [-0.21, 0.23], $BF_{01} = 8.17$]. As predicted, observers could only integrate information from up to four out of the six available items, replicating the capacity limit observed in Gillies et al. (2023).

7.2. VWM task

The linear regression between VWM capacity and subset size revealed that participants remembered more information as set size increased, $r^2 = 0.46$, $p < .001$ (see Fig. 3B). Normality was violated for one of the planned comparison *t*-tests, so reported here is the non-parametric equivalent (Wilcoxon Signed Rank Test). Planned comparisons revealed that participants remembered significantly more information from set size one to two [$Z\,(81) = 7.86$, $p < .001$, rank-biserial correlation = 1, $BF_{10} = 619{,}989.91$], and from set size two to four [$t\,(81) = 9.79$, $p < .001$, Cohen's $d = 1.08$, *95% CI* [0.81, 1.35], $BF_{10} = 3.09 \times 10^{12}$]. However, there was no significant difference in capacity from set size four to six [$t\,(81) = 1.93$, $p = .06$, Cohen's $d = 0.21$, *95% CI* [-0.43, 0.01], $BF_{01} = 1.41$].

This pattern, which is strikingly similar to that observed in the ensemble task (but note that the BF for the set size 4–6 comparison does not indicate substantial evidence in favor of the null hypothesis), suggests that VWM capacity may explain the capacity limit observed in cross-modal ensemble taste perception. To provide more definitive support for this notion, we split individuals into "high capacity" and "low capacity" groups. If VWM resources explain the capacity limit encountered in the average taste ensemble task, then individuals with high VWM capacities should have greater "ensemble abilities" than low-capacity individuals, and hence will be able to encode and maintain more food pictures in VWM, translating into more items integrated into their ensemble percepts.

7.3. VWM split

Despite being restricted by a low trial number for the split half analyses, split half reliability for VWM capacity was significant, $r^2 = 0.06$, $p = .02$. We also applied the Spearman-Browne formula to correct for the noise ceiling. Using this formula, the $r^2 = 0.15$. For this analysis, participants with an average VWM capacity greater than 2.61 were considered "high capacity" (N = 43), and those with a capacity of less than 2.61 were considered "low capacity" (N = 39). Despite low-capacity individuals having significantly lower "VWM ability" compared to high-capacity individuals [$t\,(80) = 11.72$, $p < .001$, Cohen's $d = 2.59$, *95% CI* [1.99, 3.18], $BF_{10} = 6.83 \times 10^{15}$], there was no difference in "ensemble ability" between high and low-capacity individuals [$t\,(80) = 0.34$, $p = .73$, Cohen's $d = 0.08$, *95% CI* [-0.36, 0.51], $BF_{01} = 4.13$] (see Fig. 4). This suggests that VWM capacity was not related to participants' performance on the ensemble task, as both low and high-capacity individuals had remarkably similar ensemble abilities.

8. Experiment 2

The results of Experiment 1 suggest that differences in VWM capacity do not explain performance on the ensemble task. However, there is a possibility that participants were not relying on item-specific memory in the VWM task. Rather, the task could potentially be accomplished by using simple item features (e.g., color or shape). For example, if a participant was shown an array of foods that contained an egg and were then shown an egg and a cucumber (the distractor), they could get the answer correct by recognizing that they were shown something yellow, even if they were not certain that what they were shown was an egg. As such, performance on the VWM task may not reveal if participants have explicit memory for the items that were shown to them in the ensembles.

Experiment 2 circumvents this potential issue by using a free recall task rather than a 2AFC task. After being shown the ensemble arrays, participants were asked to list the food items they remembered seeing.

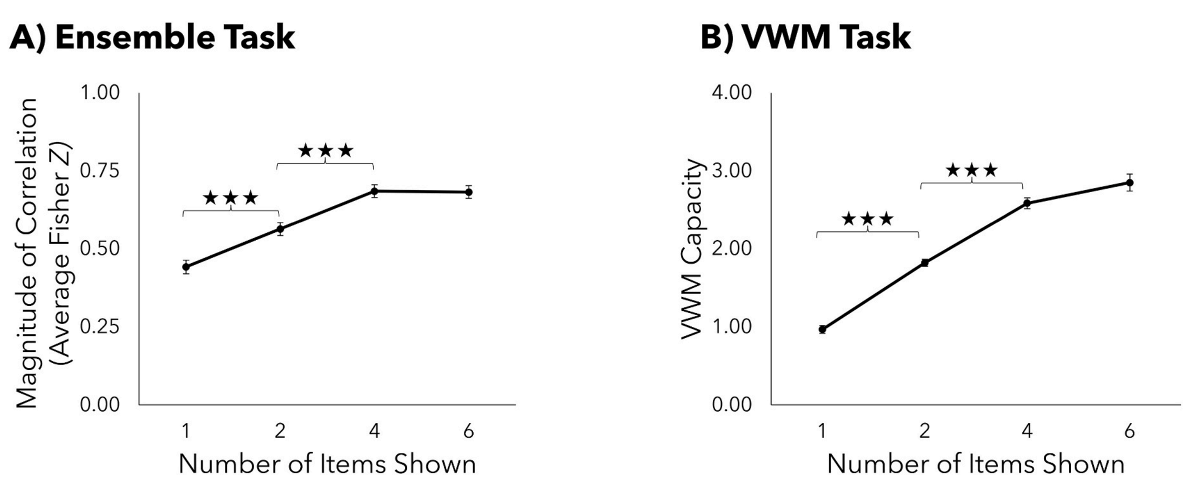

Fig. 3. Results for Experiment 1
A) The results of the ensemble task. The Y axis represents the average of the participants Fisher *Z* scores for that particular set size condition. The magnitude of the correlations increased with set-size but peaked at a maximum of four items. This provides evidence that observers were limited in the number of items they could incorporate into their cross-modal ensemble percepts.
B) The results of the VWM task. VWM capacity increased with the number of items shown, and then plateaued after four items. Error bars represent Morey's Standard Error of the Mean (SEM) (Morey, 2008). ***$p < .001$.

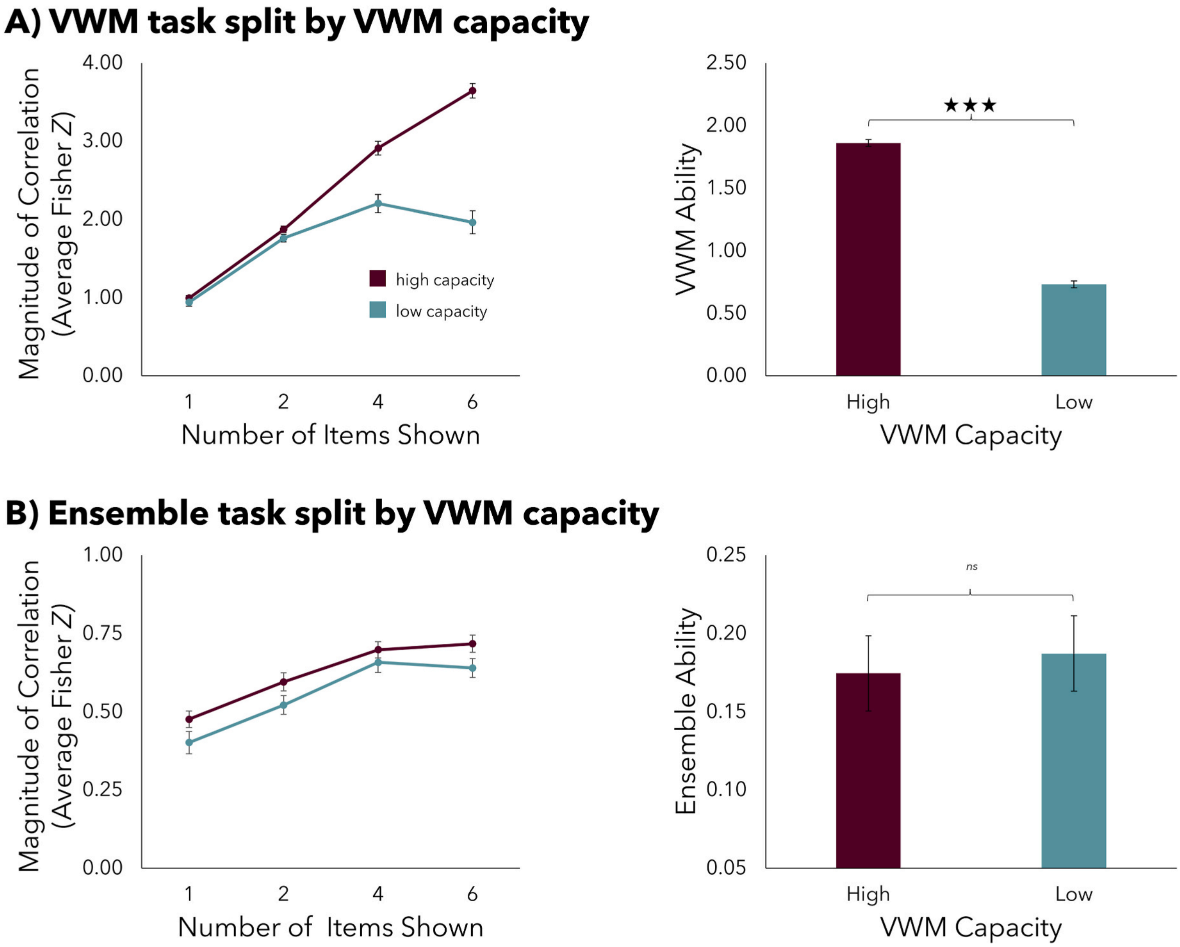

Fig. 4. Results for Experiment 1: split by VWM capacity
A) VWM task performance split by high- and low-VWM capacity individuals. The plot on the left shows how VWM capacity changes with subset size. The plot on the right shows how VWM ability (the average VWM capacity for subset 1 and 2 subtracted from the average VWM capacity for subset 4 and 6) differs by VWM capacity (high or low). Participants with low VWM capacity had significantly lower VWM ability than those with high VWM capacities.
B) Ensemble task performance split by high- and low-VWM capacity individuals. The plot on the left shows how the magnitude of the correlation changes with subset size. The plot on the right shows how ensemble ability (the average Fisher *Z* for subset size 1 and 2 subtracted from the average Fisher *Z* for subset 4 and 6) differs based on VWM capacity (high or low). Despite a difference in VWM ability, participants did not significantly differ in their ensemble abilities.
Error bars represent Morey's SEM (Morey, 2008). ***$p < .001$; *ns* = not significant.

This task enabled us to determine what items participants had explicit memory for from the preceding array. To preview, we find a pattern of results identical to that of Experiment 1.

9. Methods

9.1. Participants

Participants were recruited from Prolific (2021) using the same pre-screening procedure and payment details as in Experiment 1. Participants who completed Experiment 1 were not permitted to participate in Experiment 2.

A total of 92 participants were recruited, and eight were excluded (see 'Participant Exclusion Criteria' in Experiment 1 for details), leading to a final sample of 84. (The target for the final sample was 82, but an additional two people were accidently permitted to take part in the experiment. Their inclusion did not change the pattern of results.)

The mean age of the final sample was 30.85 years, with 53 females and 31 males. Eighty-one were right-handed, and three were left-handed. All had normal or corrected-to-normal vision, with 33 wearing glasses, 14 wearing contacts, and the rest with neither.

9.2. Apparatus

The apparatus was identical to that used in Experiment 1.

9.3. Stimuli

The images and ensembles were identical to those used in Experiment 1. The only difference was the replacement of the 2AFC task used in Experiment 1 with a single-item naming task and a free recall task.

9.4. Procedure

The ensemble task was identical to that used in Experiment 1.

For the single-item naming task, participants were instructed to view a single food picture on the screen and then type out the name of the food on the following screen. They were told to be specific with their answers (e.g., "granny smith apple" as opposed to "apple", or "mandarin orange" as opposed to "orange"). Participants were asked to maintain fixation on a white cross for 500ms and were then shown a single food picture for 1 s, followed by a 500ms delay where only the fixation cross was visible. Following the delay, participants indicated which food they saw using their keyboard, with no time limit. On the bottom of the

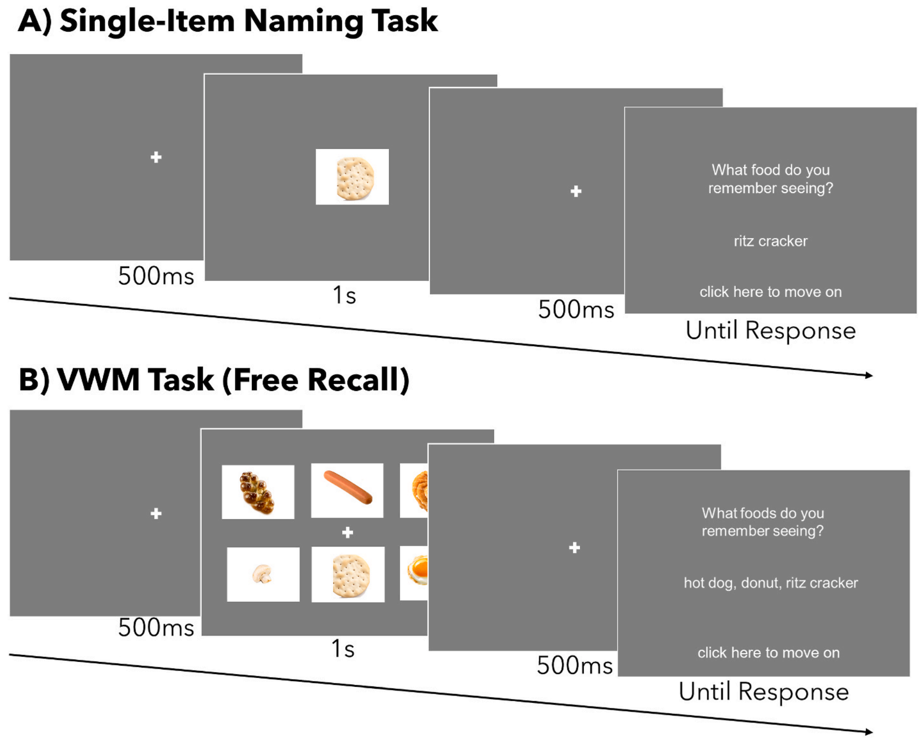

Fig. 5. Single-Item Naming and Free Recall Task Trial Sequences
A) The single-item naming task used in Experiment 2. Participants viewed one item for 1s, and then typed what they thought the item was.
B) The VWM task (free recall) used in Experiment 2. This is an example of the full six-item condition. Participants viewed the images for 1s, and were then asked to type out what items they recalled from the previous display.

screen was a clickable instruction in white font reading "click here to move on". Participants saw all 150 food pictures for a total of 150 trials (see Fig. 5).

For the free recall task, participants were instructed to remember the individual items from groups of food pictures that could vary in set size, while maintaining fixation on a central cross (which was present 500ms prior to the appearance of the ensemble and remained on screen until the response screen appeared). Following a 500ms delay, participants indicated which foods they recalled being shown using their keyboard. Participants could see their typed responses and could use the backspace key to edit a response if they made a typo. There was no time limit imposed on the response screen. On the bottom of the screen was a clickable instruction in white font reading "click here to move on". Participants saw each of the 25 ensembles at each subset size except the subset size one condition for a total of 75 trials (see Fig. 5).

Participants completed the three tasks in one of two possible task orders, 1) single-item naming task, free recall task, ensemble task; or 2) ensemble task, single-item naming task, free recall task. Task order was counterbalanced across participants.

10. Analysis

10.1. Ensemble task

Experiment 2 uses the same set of analyses as described in Experiment 1.

10.2. VWM free recall task

Four independent scorers were used to manually mark each participant's responses on the free recall task. As we could not assume participants would name every food picture in exactly the same way, their responses on the free recall task were compared to their own responses on the single-item naming task. A response was considered "correct" if the description participants gave was close to their original description in the single-item naming task. Answers were marked incorrect if the description was of a different food, was too vague, or was missing important modifiers. For example, answering "apple" in the free recall task if that same item was originally called "granny smith apple" in the single-item naming task would be marked incorrect, as its unclear what type of "apple" the participant was remembering (e.g, granny smith apples are more sour than red delicious apples). Prior to grading the files, all the scorers first marked one test file, and they were found to be in high agreement with one another (ICC = 0.96).

As a more lenient measure of VWM capacity, we also determined the number of items answered on each trial, regardless of the accuracy of the response. Our rationale for doing so was guided by the observation that the representation of an incorrectly recalled food item could still be used to generate a summary percept of average sweetness in the ensemble task. Indeed, an incorrect food representation could have a similar sweetness value compared with a correct representation (e.g., recalling "chocolate cookie" when the correct item was "chocolate donut"). In such cases, the incorrect representation could still be beneficial in the ensemble task, despite it not being accurate. To account for this, the more lenient analysis assumes that *any* item that was held in VWM may be used in the ensemble task, such that the number of items recalled across the subset sizes is more important than how precise or accurate that information is. We then determined the average number of items that were correctly recalled for each participant and set size condition (1, 2, 4, and 6). Responses on the single-item naming task were used as data for the set size one condition. As we were comparing participant responses on the free recall task to responses on the single-item naming task, all responses on the single-item naming task were counted as "correct" unless left blank. Of note, no participant gave answers that were "incorrect" (e.g., labelling a banana as a cake) during the single-item naming task. We then performed the same set of analysis as described in Experiment 1.

10.3. VWM split

The VWM split proceeded in the same way as in Experiment 1.

In addition to splitting by VWM capacity calculated from performance on the free recall task, we also repeated the analysis using the more lenient criteria of average number of items answered (where accuracy was not a factor). For this analysis, as done previously, "VWM ability" was defined as the average number of items answered at set size 1 and 2 subtracted from the average number of items answered at set size 4 and 6.

11. Results and discussion

11.1. Ensemble task

As in Experiment 1, participants incorporated more information as it was made available to them, $r^2 = 0.08$, $p < .001$ (see Fig. 6A). Planned paired-sample *t*-tests showed there was a significant increase in the magnitude of the correlation between set size one to two [$t\,(83) = 3.19$, $p < .01$, Cohen's $d = 0.35$, *95% CI* [0.12, 0.57], $BF_{10} = 12.71$], and between set size two to four [$t\,(83) = 3.81$, $p < .001$, Cohen's $d = 0.17$ *95% CI* [0.19, 0.64], $BF_{10} = 79.20$]. Finally, there was no significant difference in the magnitude of the correlation between set size four and six [$t\,(83) = 1.58$, $p = .12$, Cohen's $d = 0.17$, *95% CI* [-0.04, 0.39], $BF_{01} = 2.54$], replicating the capacity limit observed in Experiment 1 and in previous research (Gillies et al., 2023). Note that, while the Bayes Factor comparing the difference from set size four to six only provides anecdotal evidence in favor of the null hypothesis, the effect is actually in the opposite direction expected for ensemble integration (i.e., a decrease in the *Fisher Z* value from set size four to six). Therefore, there is no evidence that observers could integrate any more than four out of the six available items into their percepts of ensemble sweetness.

11.2. VWM free recall task

The linear regression revealed that participants remembered more information as set size increased, $r^2 = 0.57$, $p < .001$ (see Fig. 6B). Normality was violated for all the planned comparison *t*-tests, so reported here are the non-parametric equivalent tests (Wilcoxon Signed Rank Test). Planned comparisons revealed that participants remembered significantly more information from set size one to two [$Z\,(83) = 7.95$, $p < .001$, rank-biserial correlation $= 0.99$, $BF_{10} = 262{,}909.79$], and from set size two to four [$Z\,(83) = 7.81$, $p < .001$, rank-biserial correlation $= 0.98$, $BF_{10} = 84{,}14.28$]. Similar to Experiment 1, there was no significant increase from set size four to six [$Z\,(83) = 1.41$, $p = .26$, rank-biserial correlation $= 0.15$, $BF_{01} = 5.02$].

11.3. VWM free recall task, lenient criteria

The linear regression revealed that the number of items participants answered increased with increasing set sizes, $r^2 = 0.72$, $p < .001$ (see Fig. 6C). Normality was violated for the planned comparison *t*-tests, so reported here are the non-parametric equivalents (Wilcoxon Signed Rank Tests). There was a significant increase in the number of items answered between set size 1 and 2 [$Z\,(83) = 7.96$, $p < .001$, rank biserial correlation $= 1$, $BF_{10} = 42{,}824.47$], set size 2 and 4 [$Z\,(83) = 7.86$, $p < .001$, rank biserial correlation $= 1$, $BF_{10} = 5.22 \times 10^6$], and between set size 4 and 6 [$Z\,(83) = 3.32$, $p < .001$, rank biserial correlation $= 0.45$, $BF_{10} = 64.25$].

11.4. VWM free recall split

The split-half reliability for VWM capacity was high, $r^2 = 0.80$, $p < .001$, indicating that our measure of VWM capacity was internally consistent. Using the Spearman-Browne formula, the $r^2 = 0.90$. For this analysis, participants with an average VWM capacity greater than 1.90 were considered "high capacity" (N = 42), and those with a capacity of less than 1.90 were considered "low capacity" (N = 42).

The normality assumption was violated for one of the comparisons, so reported here is the non-parametric equivalent for that comparison (Welch's test). Despite low-capacity individuals having significantly lower "VWM ability" compared to high-capacity individuals [$t\,(49.67) = 8.53$, $p < .001$, Cohen's $d = 1.86$, *95% CI* [1.29, 2.42], $BF_{10} = 11{,}286.45$], there was no difference in "ensemble ability" between high and low-capacity individuals [$t\,(82) = 0.44$, $p = .70$, Cohen's $d = 0.10$, *95% CI* [-0.32, 0.52], $BF_{01} = 4.03$] (see Fig. 7).

This replicates the findings in Experiment 1, and together these two experiments show that VWM capacity is not related to ensemble ability, as both low and high-capacity individuals did not differ in their ensemble perception abilities.

11.5. VWM lenient grading split

The split half reliability for this measure showed that it was consistent, $r^2 = 0.43$, $p < .001$. Using the Spearman-Browne formula, the $r^2 = 0.62$.

For this analysis, participants with an average VWM capacity greater than 2.55 were considered "high capacity" ($N = 42$), and participants

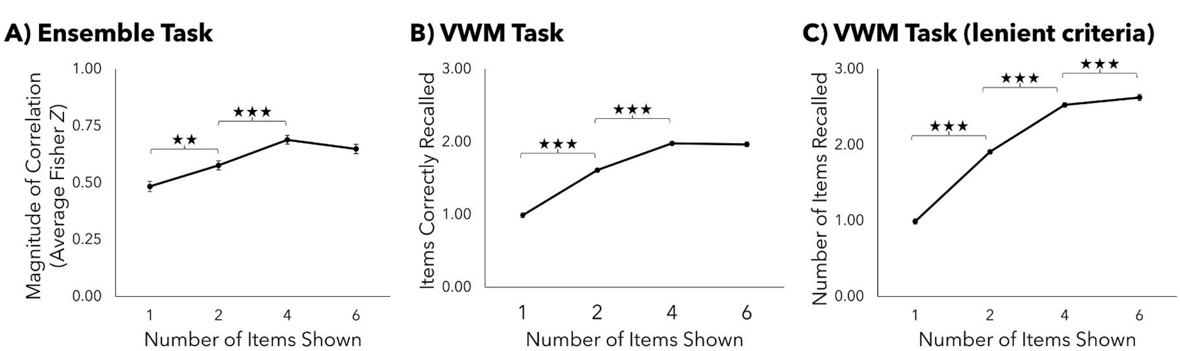

Fig. 6. Results for Experiment 2

A) The results of the ensemble task. The Y axis represents the average of the participants Fisher *Z* scores for that particular set size condition. The magnitude of the correlations increased with set-size but peaked at a maximum of four items. This provides evidence that observers were limited in the number of items they could incorporate into their cross-modal ensemble percepts.

B) The results of the VWM task. VWM capacity increased with the number of items shown, and then plateaued after four items.

C) The results of the VWM task with lenient criteria. Capacity increased with the number of items shown but did not plateau (likely because accuracy was not taken into account).

Error bars represent Morey's SEM (Morey, 2008). ***p* < .01; ****p* < .001.

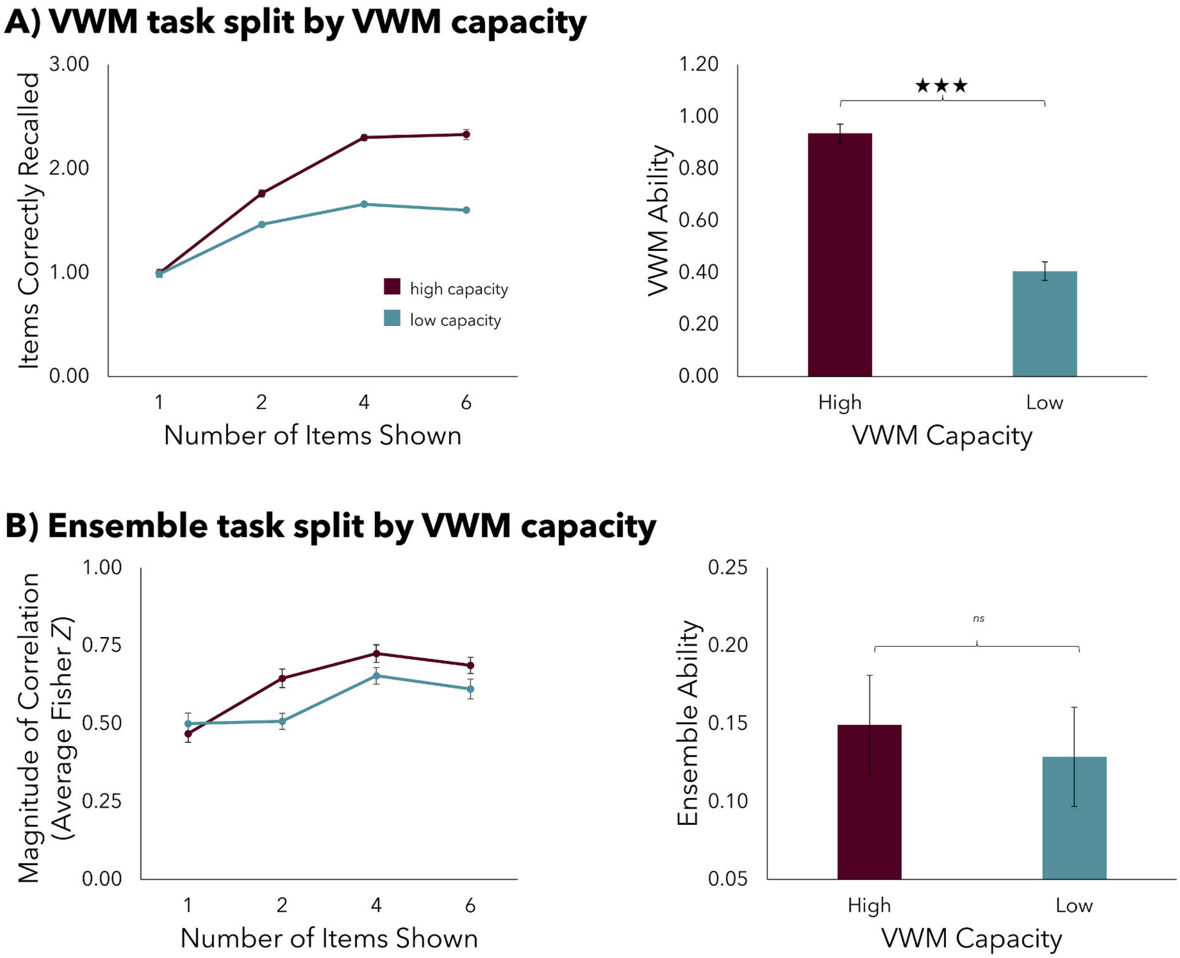

Fig. 7. Results for Experiment 2: split by VWM capacity
A) VWM task performance split by high- and low-VWM capacity individuals. The plot on the left shows how VWM capacity changes with subset size. The plot on the right shows how VWM ability (the average VWM capacity for subset 1 and 2 subtracted from the average VWM capacity for subset 4 and 6) differs by VWM capacity (high or low). Participants with low VWM capacity had significantly lower VWM ability than those with high VWM capacities.
B) Ensemble task performance split by high- and low-VWM capacity individuals. The plot on the left shows how the magnitude of the correlation changes with subset size. The plot on the right shows how ensemble ability (the average Fisher *Z* for subset size 1 and 2 subtracted from the average Fisher *Z* for subset 4 and 6) differs based on VWM capacity (high or low). Despite a difference in VWM ability, participants did not significantly differ on their ensemble abilities.
Error bars represent Morey's SEM (Morey, 2008). $^{***}p < .001$; *ns* = not significant.

with an average VWM capacity less than 2.55 were considered "low capacity" ($N = 42$). The normality assumption was violated for one of the comparisons, so reported here is the non-parametric equivalent for that comparison (Welch's test).

Despite a significant difference in VWM ability between low and high capacity individuals [$t\,(62.26) = 9.42$, $p < .001$, Cohen's $d = 2.17$, *95% CI* = [1.59, 2.74], $BF_{10} = 33{,}095.64$], there was no significant difference in ensemble ability between these two groups [$t\,(82) = 1.00$, $p = .32$, Cohen's $d = 0.22$, *95% CI* = [-0.21, 0.65], $BF_{01} = 2.83$] (see Fig. 8). Therefore, even when ignoring accuracy in the free recall task with the use of a very lenient metric to define VWM capacity, there was no clear relationship between how many items people recalled and their performance on the ensemble task. Together with the free recall results and the results of Experiment 1, this suggests that VWM capacity does not adequately explain the capacity limitation observed in the cross-modal perception of ensemble taste.

12. General discussion

Across two experiments, we showed that participants were limited in their ability to form cross-modal ensemble percepts for taste from visually presented information. As shown previously (Gillies et al., 2023), observers could only integrate taste information from four out of six available food items. Here we examined if this capacity limitation in ensemble encoding could be explained by VWM resources. Interestingly, although the pattern of VWM performance mirrored performance on the ensemble task (i.e., an increase in the number of items integrated until a plateau at subset size four on the ensemble task, and an increase in the number of items remembered until a plateau at subset size four on the VWM task in both Experiment 1 and 2), we subsequently found that there was no relationship between VWM capacity and cross-modal ensemble perception of taste. Specifically, individuals with higher VWM capacities did not perform differently on the ensemble task than low-capacity individuals. This lack of a relationship between VWM and ensemble coding was found when we defined VWM capacity using both a 2AFC recognition task, and a more nuanced free recall task. Although our data cannot definitively state whether or not VWM is *necessary* to perform the ensemble task, it can certainly state that it is not *sufficient*.

13. Did we adequately and reliably measure VWM?

One possible reason we did not find a relationship between VWM and ensemble coding may be that our measure of VWM was inadequate. In Experiment 1, observers did not need to hold item-level information in

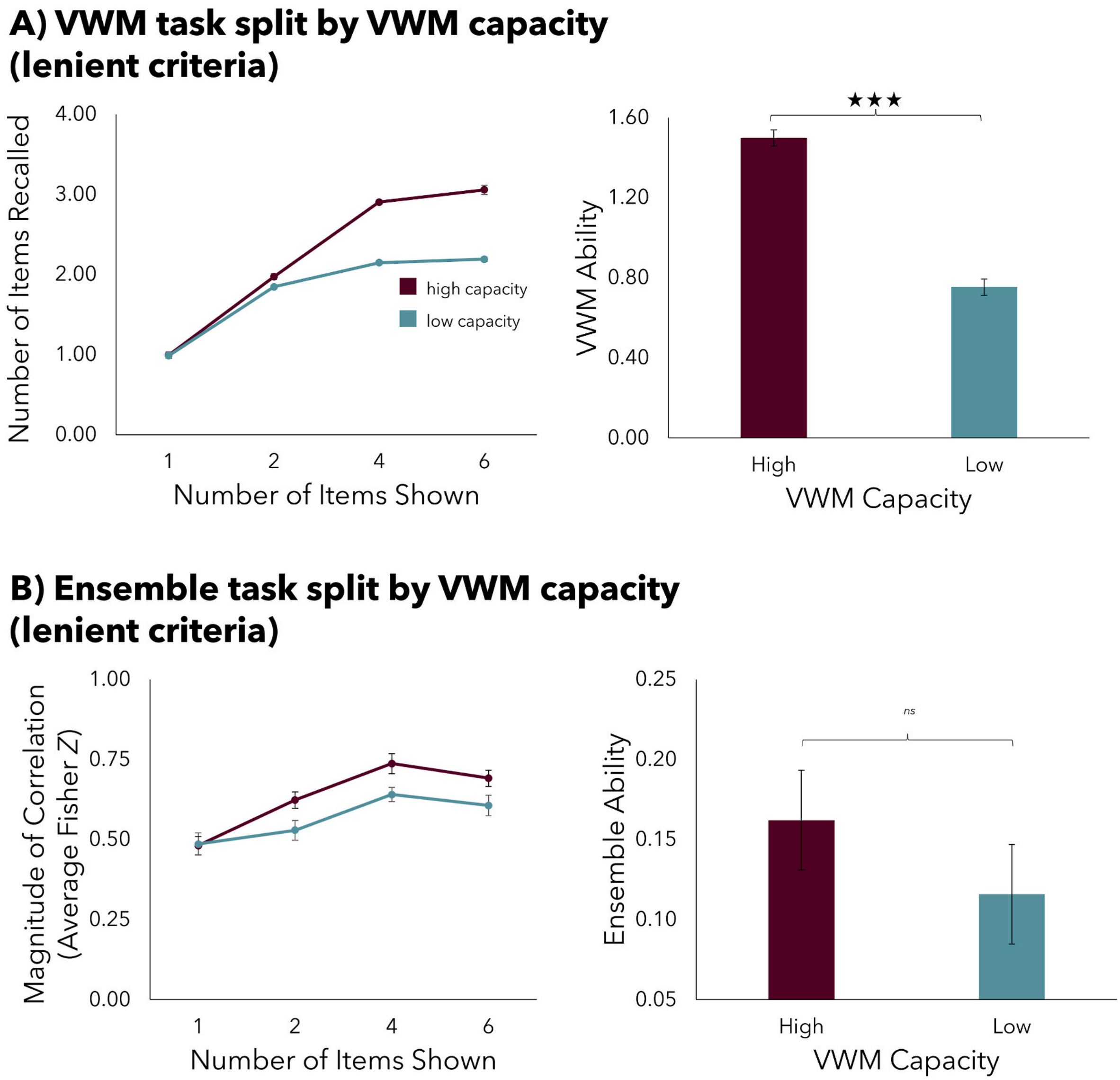

Fig. 8. Results for Experiment 2: split by VWM capacity (lenient criteria)
A) VWM task performance split by high- and low-VWM capacity individuals. The plot on the left shows how VWM capacity changes with subset size. The plot on the right shows how VWM ability (the average VWM capacity for subset 1 and 2 subtracted from the average VWM capacity for subset 4 and 6) differs by VWM capacity (high or low). Participants with low VWM capacity had significantly lower VWM ability than those with high VWM capacities.
B) Ensemble task performance split by high- and low-VWM capacity individuals. The plot on the left shows how the magnitude of the correlation changes with subset size. The plot on the right shows how ensemble ability (the average Fisher *Z* for subset size 1 and 2 subtracted from the average Fisher *Z* for subset 4 and 6) differs based on VWM capacity (high or low). Despite a difference in VWM ability, participants did not significantly differ on their ensemble abilities.
Error bars represent Morey's SEM (Morey, 2008). $^{***}p < .001$; *ns* = not significant.

VWM to perform above chance on the 2AFC task. Rather, participants could have used low-level feature information (e.g., color, shape) to perform the 2AFC task. Importantly, these low-level features do not contribute useful information about the taste of food pictures. While some visual features may contribute to taste perception (Spence et al., 2010; Spence, 2015) (e.g., red strawberries are sweeter than pale ones), information conveyed by these features is not consistent across different food items (e.g., red foods are not always sweet). Therefore, even if low-level feature information was used in the VWM task, this information would not be helpful to participants in the ensemble task.

However, even if we concede that it is possible that observers are relying on low-level features in Experiment 1 to perform the VWM task, this strategy was not possible in Experiment 2. Experiment 2 used a free recall task, meaning that observers could not rely on low-level visual features to recall items they had seen in the previous food ensemble. Rather, participants would need to retrieve item-specific information from VWM. As the pattern of results was identical between Experiment 1 and Experiment 2, which used different tasks to define VWM capacity, we can confidently conclude that we did indeed adequately measure VWM, and such resources are not sufficient to explain the capacity limitation observed in the cross-modal perception of ensemble taste.

Another point to consider is that our measure of VWM capacity may not reliably measure individual differences (Hedge et al., 2018). However, split-half reliability for the VWM capacity measure was significant for both Experiment 1 and 2. Split-half reliability was lower in Experiment 1, but the number of trials was lower than other studies that used indirect measures of VWM capacity (e.g., Xu et al., 2018). Yet despite being limited by the number of trials, split-half reliability was still significant in Experiment 1. Taken together, we demonstrate good split-half reliability across both measures of VWM, and thus conclude that we were able to reliably examine individual differences in VWM capacity in both Experiments 1 and 2. Given that the significant differences in the number of items recalled between low- and high-capacity individuals did not translate into a significant difference in ensemble perception between the groups, we again conclude that the capacity limitation observed in ensemble taste perception cannot be fully explained by differences in VWM resources across participants.

14. Other instances where VWM does not contribute to ensemble coding

Our finding that VWM capacity differences did not fully explain ensemble coding abilities is in line with previous studies. Yamanashi-Leib et al. (2016, 2020) found that observers were sensitive to abstract features of groups of objects (i.e., average animacy, average cost). Animacy and cost cannot be computed using information from a single feature (e.g., color, size), rather, the perception of animacy or cost likely arises from the interactions of several features, with cost likely relying more on non-visual semantic information. Utilizing the ensemble task design used in the present work, they found that observers could integrate information from all the available objects in an ensemble (i.e., 6 objects), superseding their measured VWM capacities. Furthermore, observers had poor memory for individual items relative to their ensemble coding abilities, as participants could not remember all the items that they were nonetheless able to incorporate into their ensemble percepts.

In the above studies, it is clear that VWM alone cannot explain performance on the ensemble coding tasks. These results are consistent with what we find regarding the relationship between VWM and ensemble taste perception in the present study. However, these previous studies also observed seemingly capacity-unlimited ensemble coding abilities, which is in contrast to the capacity limited results we observe when observers are required to extract summary information of taste from visually presented food items. While a capacity limitation of about four ensemble items seems qualitatively similar to the classic finding of a VWM capacity limit (Luck and Vogel, 1997), subsequent analyses (see Figs. 4, 7 and 8) revealed that VWM resources do not explain this ensemble capacity limitation. This raises the intriguing question of which cognitive processes are driving this capacity limitation.

15. Which cognitive processes explain the ensemble capacity limit?

Presently, it is unclear if this capacity limit generalizes to all cross-modal ensemble coding, or if it is something that is specific to the interaction between taste and vision. To our knowledge, no other studies examining interactions between different sensory modalities and ensemble coding have been published. Thus, further studies are necessary to determine if this generalizes to other forms of cross-modal ensemble processing (e.g., can observers extract average weight, which is initially derived from tactile information, from visually presented object pictures?).

One possible explanation for the capacity limit is that it is not a limit at all, but a byproduct of observers engaging in a purposeful subsampling strategy. Subsampling can be used to produce accurate summary statistics (Marchant et al., 2013; Lau and Brady, 2018). However, our study was designed in such a way that subsampling cannot be a viable strategy, as using only a subset of the available information produces inaccurate summary statistics. Therefore, a subsampling strategy here is not optimal and would leave observers at a disadvantage. The only optimal strategy to produce an accurate summary statistic in the current paradigm is to incorporate information from all six items. Other studies that use both complex stimuli and abstract ensemble tasks (e.g., average lifelikeness, average economic value; Yamanashi-Leib and colleagues, 2016 & 2020) found that observers did not have to rely on a subsampling strategy, as they were able to incorporate information from all the available items. Given this, we think it unlikely that subsampling items from the ensemble display explains the capacity limitation we observe here and in our previous study (Gillies et al., 2023).

How do observers determine sweetness from visually presented information? Some visual features can contribute to taste perception (Spence et al., 2010; Spence, 2015) (e.g., red strawberries are sweeter than pale ones). However, the information conveyed by these visual features is not consistent (e.g., red foods are not always sweet). As such, the visual features associated with a food's taste vary across food categories such that an ensemble value for multiple foods would not be related to the shared visual features of those foods. While it is unclear precisely what type of process observers use to extract stored taste values, it is plausible that observers *did* rely on their VWM to hold item-level information, and then used this information to aide in retrieving representations of sweetness associated with those items. However, given the lack of a relationship between VWM and high-level ensemble coding observed here (and in previous studies; Yamanashi-Leib et al., 2016, 2020), the locus of the bottleneck that produces the capacity limit is not likely to be at the stage of encoding individual items into VWM. Given this, we would like to highlight several possible candidates for the source of the bottleneck.

One possible candidate relates to a bottleneck at the perceptual stage. More specifically, the capacity limit may emerge as a failure of parallel visual processing of multiple features across many food items, as there are no visual features that convey sweetness on their own. Parallel processing mechanisms may encounter a difficulty in pooling sensory signals across the display, and observers would in this case be forced to subsample from the display. As such, parallel processing may constrain the number of visual cues that an observer can integrate at a time to trigger the necessary cross-modal retrieval of sweetness information that is needed to compute an average sweetness.

Alternatively, another possible candidate relates to the interaction of multiple working memory systems. Specifically, we speculate that participants may: 1) encode item-level information into VWM, and then 2) use that information to guide the retrieval of stored taste information, and finally 3) hold that taste information within another working memory system, specifically gustatory working memory (GWM). We contend that it is possible that the capacity limitation observed in ensemble taste perception may be due to a bottleneck encountered within GWM (see Fig. 9). Indeed, while only a few studies have investigated the idea of a short-term storage system for taste, the little work that has been done suggests that GWM is limited to about three tastes (Lim et al., 2022; Daniel and Katz, 2018).

To test the capacity of GWM, Lim et al. (2022) had participants taste colorless, odorless liquids that were either sweet, salty, sour, or bitter. Tastes were presented in sequence at differing set sizes (i.e., one to five tastes). Participants were then presented with a probe taste, and they indicated if that taste was present in the previous sequence of tastes. Accuracy on the probe task decreased as set size increased, approaching chance performance around set size three. While it is quite understudied, the existing research converges on the idea that GWM is capacity limited, like other working memory systems. The crucial point here is that the operation of a limited capacity working memory system on its own (be it visual or gustatory) may not be sufficient to explain the capacity limitation observed in ensemble taste perception. Instead, it could be the coordinated interaction between distinct working memory systems (i.e., visual and gustatory) that leads to a bottleneck in processing and manifests as an ensemble capacity limitation.

Just holding the visual information in VWM is not a guarantee that it can be successfully converted to taste information in GWM. Studies have found that working memory systems are multiplexed, with dissociable storage mechanisms for separate sensory information (e.g., Crottaz-Herbette et al., 2004; Fougnie et al., 2015; Fougnie and Marois, 2011; Katus and Eimer, 2018; Cocchini et al., 2002). Interestingly, the capacity for the sensory storages is not unanimous within individuals, and thus having a large capacity in one sensory working memory system does not guarantee an equally large capacity in another sensory working memory modality. This suggests that, even if an individual has a large VWM

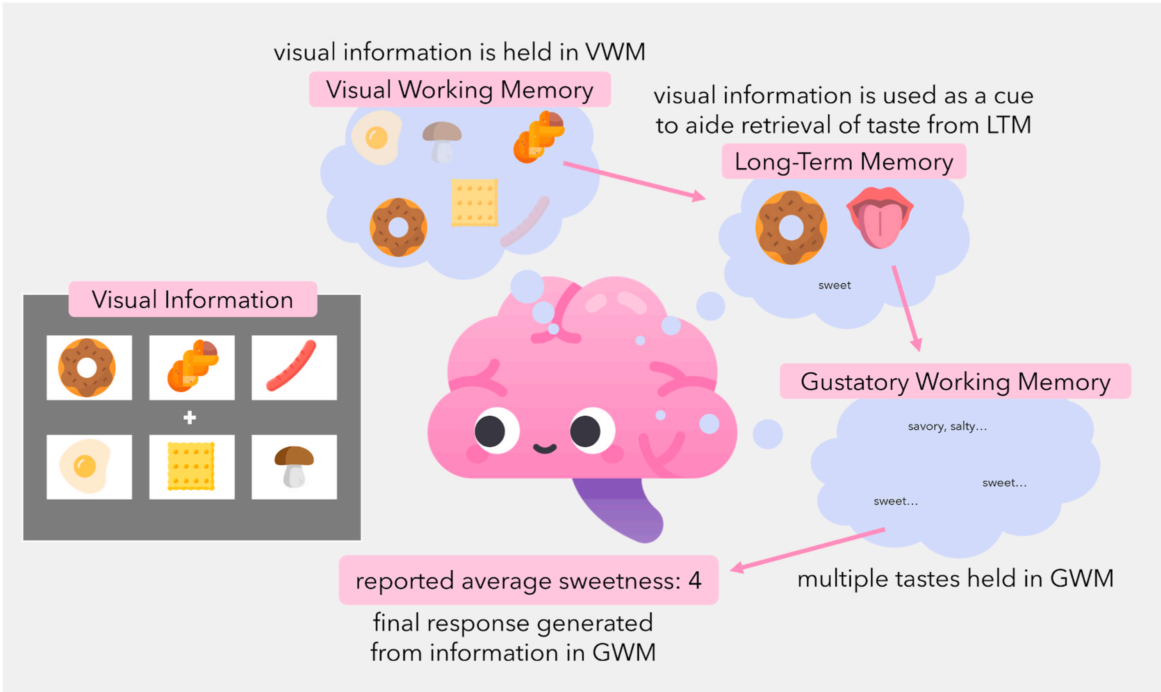

Fig. 9. A Proposed Mechanism of Vision-Taste Ensemble Coding
An illustration of how average sweetness information may be generated. First, observers may hold some of the information in VWM with varying levels of fidelity/precision across the items (the first bottleneck). Then, the visual information is used to cue retrieval of taste from LTM. Next, multiple tastes are held within GWM. Another bottleneck may occur at these latter two stages, as participants may fail to retrieve all the tastes from LTM, or they may not be able to hold all the tastes in GWM. Last, the average sweetness score is generated using the information held in GWM. Here, VWM alone cannot explain performance on the ensemble sweetness task.

capacity, this individual may very well have a small GWM capacity.

Another possibility to consider is the potential involvement of verbal working memory. In the free recall task used in Experiment 2, rather than holding the information in VWM, participants may have attempted to label each item using verbal working memory. Indeed, we did ask participants to produce written answers. Research has shown that working memory is modality specific (Crottaz-Herbette et al., 2004; Fougnie et al., 2015; Fougnie and Marois, 2011; Katus and Eimer, 2018; Cocchini et al., 2002), and we cannot assume that verbal and visual working memory ability correlate perfectly with one another. This leads to the question of whether or not participants use their verbal working memory during ensemble encoding, and whether a bottleneck in the former system explains the capacity limitation we observe in the perception of ensemble taste. To examine the contribution of verbal working memory, a future study can incorporate an articulatory suppression task (Baddeley et al., 1975) during the ensemble and free recall phases of the experiment. Of note, we are not suggesting that participants used either verbal *or* visual working memory on their own, but that they may have engaged in a number of different cognitive strategies. In this case, it is possible that multiple different systems interact during ensemble coding.

If GWM or verbal working memory is the limiting factor, then there should be a relationship between individuals' modality-specific working memory resources and their ability to produce a summary statistic for taste from visual information. This could be the case whether or not GWM or verbal working memory performance predicts VWM performance (i.e., individuals with high VWM ability may or may not also have high GWM and/or verbal working memory abilities). Alternatively, as alluded to above, it could be the interactive operation of all three working memory systems that explains the capacity limitation observed in ensemble taste processing. Another possibility is that the bottleneck may not be due to the interaction between working memory systems themselves, but instead on the utilization of their shared contents to trigger the retrieval of information from LTM. Further research is needed to examine these possibilities, which could entail investigating gustatory, verbal, and visual working memory simultaneously in a study examining ensemble encoding of average taste from real-life food items. Doing so will not only shed light on the cognitive processes underlying ensemble taste perception, but also on the mechanics of how different cognitive systems (ensemble coding, VWM, GWM, verbal working memory, LTM) interact in the production of everyday behavior.

16. Distinct systems can still interact

The finding of independence between different neural systems does not suggest that these systems do not interact in the production of skilled behavior. Take, for example, the operation of the dorsal and ventral streams of visual processing (Goodale and Milner, 1992), which mediate vision-for-action and vision-for-perception (Goodale and Milner, 2013), respectively. Numerous behavioral (e.g., Cant et al., 2005; Hu and Goodale, 2000), neuroimaging (e.g., Culham et al., 2003; Valyear et al., 2006), and, particularly, neuropsychological (e.g., Goodale et al., 1991; James et al., 2003) research has demonstrated that these cortical processing streams can function independently of one another. However, despite this independence, the streams are not hermetically sealed from each other (Goodale, 2011), and constantly interact in everyday life. Indeed, some actions programmed by the dorsal stream rely on input from the ventral stream for the successful execution of visuomotor behavior. For example, when interacting with objects made of different materials (e.g., light objects made of polystyrene, or heavy objects made of metal), perceptual processing of the surface and material properties of objects carried out by regions in the ventral stream (Cant and Goodale, 2007, 2011) is communicated to visuomotor regions in the dorsal stream to inform the initial grip and load forces needed to successfully pick up the objects without having them slip out of your fingers (Buckingham et al., 2009; Gallivan et al., 2014).

Translating this line of reasoning to the present study, the finding that ensemble perception of average taste is distinct from the operation of VWM does not suggest that these two cognitive systems do not interact whatsoever. Similarly, the finding that ensemble perception does not influence visuomotor behavior (i.e., reaching and grasping) does not imply that actions cannot be influenced by the perception of summary statistics (e.g., ensemble perception may contribute to memory-guided, as opposed to visually-guided, behavior; Fan et al., 2021). Future studies are needed to determine when and how different cognitive systems may interact during ensemble coding.

17. Conclusion

In summary, we found that observers were limited in their ability to extract average taste (i.e., sweetness) from visually presented food information. Specifically, they were limited in the number of items they could incorporate into their cross-modal ensemble percepts. Interestingly, individual differences in participants' VWM capacities did not explain performance on the ensemble taste task, showing that VWM alone is not sufficient to explain the capacity limit we observed. Instead, we speculate that this limit may be related to the interaction between multiple working memory systems (VWM, GWM, verbal working memory), or in using the output of this interactive processing to retrieve cross-modal representations of taste stored in LTM. Given the pervasiveness of ensemble processing (Corbett et al., 2023), it is highly likely that the perception of summary statistics influences, and is influenced by, several cognitive, motor, and decision-related neural processes and it is the task of future studies to determine the boundary conditions underlying such interactive processing.

Author note

All authors designed the research; Greer Gillies performed research and data analysis, all authors wrote the first and subsequent drafts of the article. The authors Jonathan S. Cant and Keisuke Fukuda acquired funding from NSERC.

This work was funded by the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant (435647) to Jonathan S. Cant, and an NSERC Discovery Grant (5009170) to Keisuke Fukuda.

We would like to thank the Cant and Fukuda labs, with special thanks to Kristina Knox and the rest of the 2020 University of Toronto Psychology Graduate Cohort.

Data and analysis scripts are freely available on Open Science Framework at https://osf.io/q89un/.

Part of this work was presented at the Object Perception, visual Attention, and Memory (OPAM) conference in 2021 and the Vision Sciences Society (VSS) conference in 2022.

We have no conflicts of interest to report.

CRediT authorship contribution statement

Greer Gillies: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing - original draft, Writing - review & editing. **Keisuke Fukuda:** Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, Writing - review & editing. **Jonathan S. Cant:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing - review & editing.

Data availability

Data will be made available on request.

References

Adobe Photoshop CS, 2004. Peachpit Press, Berkeley, CA.

Baddeley, A.D., Thomson, N., Buchanan, M., 1975. Word length and the structure of short-term memory. J. Verb. Learn. Verb. Behav. 14 (6), 575–589. https://doi:10.1016/S0022-5371(75)80045-4.

Blechert, J., Lender, A., Polk, S., Busch, N.A., Ohla, K., 2019. Food-picsextended—an image database for experimental research on eating and appetite: additional images, normative ratings and an updated review. Front. Psychol. 10, 307. https://doi.org/10.3389/fpsyg.2019.00307.

Brady, T.F., Alvarez, G.A., 2011. Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. Psychol. Sci. 22 (3), 384–392. https://doi.org/10.1177/0956797610397956.

Brand, J., Oriet, C., Sykes Tottenham, L., 2012. Size and emotion averaging: costs of dividing attention after all. Canadian J. Experimental Psychol./Revue canadienne de psychologie expérimentale 66 (1), 63. https://doi.org/10.1037/a0026950.

Buckingham, G., Cant, J.S., Goodale, M.A., 2009. Living in a material world: how visual cues to material properties affect the way that we lift objects and perceive their weight. J. Neurophysiol. 102 (6), 3111–3118. https://doi.org/10.1152/jn.00515.2009.

Cant, J.S., Goodale, M.A., 2007. Attention to form or surface properties modulates different regions of human occipitotemporal cortex. Cerebr. Cortex 17 (3), 713–731. https://doi.org/10.1093/cercor/bhk022.

Cant, J.S., Goodale, M.A., 2011. Scratching beneath the surface: new insights into the functional properties of the lateral occipital area and parahippocampal place area. J. Neurosci. 31 (22), 8248–8258. https://doi.org/10.1523/JNEUROSCI.6113-10.2011.

Cant, J.S., Westwood, D.A., Valyear, K.F., Goodale, M.A., 2005. No evidence for visuomotor priming in a visually guided action task. Neuropsychologia 43 (2), 216–226. https://doi.org/10.1016/j.neuropsychologia.2004.11.008.

Cicchetti, D.V., 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol. Assess. 6 (4), 284. https://doi.org/10.1037/1040-3590.6.4.284.

Cocchini, G., Logie, R.H., Sala, S.D., MacPherson, S.E., Baddeley, A.D., 2002. Concurrent performance of two memory tasks: evidence for domain-specific working memory systems. Mem. Cognit. 30 (7), 1086–1095. https://doi.org/10.3758/BF03194326.

Cohen, M.A., Dennett, D.C., Kanwisher, N., 2016. What is the bandwidth of perceptual experience? Trends Cognit. Sci. 20 (5), 324–335. https://doi.org/10.1016/j.tics.2016.03.006.

Corbett, J.E., Utochkin, I., Hochstein, S., 2023. The Pervasiveness of Ensemble Perception: Not Just Your Average Review. Cambridge University Press. https://doi.org/10.1017/9781009222716.

Corbin, J.C., Crawford, L.E., 2018. Biased by the group: memory for an emotional expression biases towards the ensemble. Collabra: Psychology 4 (1). https://doi.org/10.1525/collabra.186.

Crottaz-Herbette, S., Anagnoson, R.T., Menon, V., 2004. Modality effects in verbal working memory: differential prefrontal and parietal responses to auditory and visual stimuli. Neuroimage 21 (1), 340–351. https://doi.org/10.1016/j.neuroimage.2003.09.019.

Culham, J.C., Danckert, S.L., Souza, J.F.D., Gati, J.S., Menon, R.S., Goodale, M.A., 2003. Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. Exp. Brain Res. 153, 180–189. https://doi.org/10.1007/s00221-003-1591-5.

Daniel, T.A., Katz, J.S., 2018. Primacy and recency effects for taste. J. Exp. Psychol. Learn. Mem. Cognit. 44 (3), 399. https://doi.org/10.1037/xlm0000437.

Fan, A.W.Y., Guo, L.L., Frost, A., Whitwell, R.L., Niemeier, M., Cant, J.S., 2021. Grasping of real-world objects is not biased by ensemble perception. Front. Psychol. 1158 https://doi.org/10.3389/fpsyg.2021.597691.

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. Behav. Res. Methods 39 (2), 175–191. https://doi.org/10.3758/BF03193146.

Foroni, F., Pergola, G., Argiris, G., Rumiati, R.I., 2013. The FoodCast research image database (FRIDa). Front. Hum. Neurosci. 7, 51. https://doi.org/10.3389/fnhum.2013.00051.

Fougnie, D., Marois, R., 2011. What limits working memory capacity? Evidence for modality-specific sources to the simultaneous storage of visual and auditory arrays. J. Exp. Psychol. Learn. Mem. Cognit. 37 (6), 1329. https://doi.org/10.1037/a0024834.

Fougnie, D., Zughni, S., Godwin, D., Marois, R., 2015. Working memory storage is intrinsically domain specific. J. Exp. Psychol. Gen. 144 (1), 30. https://doi.org/10.1037/a0038211.

Gallivan, J.P., Cant, J.S., Goodale, M.A., Flanagan, J.R., 2014. Representation of object weight in human ventral visual cortex. Curr. Biol. 24 (16), 1866–1873. https://doi.org/10.1016/j.cub.2014.06.046.

Gillies, G., Fukuda, K., Cant, J.S., 2023. Testing the flexibility of ensemble coding: limitations in cross-modal ensemble perception. J. Exp. Psychol.: General. Advance online publication. https://doi.org/10.1037/xge0001470.

Goodale, M.A., 2011. Transforming vision into action. Vis. Res. 51 (13), 1567–1587. https://doi.org/10.1016/j.visres.2010.07.027.

Goodale, M.A., Milner, A.D., 1992. Separate visual pathways for perception and action. Trends Neurosci. 15 (1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8.

Goodale, M.A., Milner, A.D., Jakobson, L.S., Carey, D.P., 1991. A neurological dissociation between perceiving objects and grasping them. Nature 349 (6305), 154–156. https://doi.org/10.1038/349154a0.

Goodale, M., Milner, D., 2013. Sight Unseen: an Exploration of Conscious and Unconscious Vision. OUP Oxford. https://doi.org/10.1002/col.20112.

Griffiths, S., Rhodes, G., Jeffery, L., Palermo, R., Neumann, M.F., 2018. The average facial expression of a crowd influences impressions of individual expressions. J. Exp. Psychol. Hum. Percept. Perform. 44 (2), 311. https://doi.org/10.1037/xhp0000446.

Hedge, C., Powell, G., Sumner, P., 2018. The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. Behav. Res. Methods 50, 1166–1186. https://doi.org/10.3758/s13428-017-0935-1.
Hu, Y., Goodale, M.A., 2000. Grasping after a delay shifts size-scaling from absolute to relative metrics. J. Cognit. Neurosci. 12 (5), 856–868. https://doi.org/10.1162/089892900562462.
James, T.W., Culham, J., Humphrey, G.K., Milner, A.D., Goodale, M.A., 2003. Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. Brain 126 (11), 2463–2475. https://doi.org/10.1093/brain/awg248.
Jeffreys, H., 1961. Theory of Probability. Oxford University Press, Oxford, England.
Ji, L., Chen, W., Loeys, T., Pourtois, G., 2018. Ensemble representation for multiple facial expressions: evidence for a capacity limited perceptual process. J. Vis. 18 (3), 17. https://doi.org/10.1167/18.3.17, 17.
Katus, T., Eimer, M., 2018. Independent attention mechanisms control the activation of tactile and visual working memory representations. J. Cognit. Neurosci. 30 (5), 644–655. https://doi.org/10.1162/jocn_a_01239.
Khayat, N., Hochstein, S., 2019. Relating categorization to set summary statistics perception. Atten. Percept. Psychophys. 81 (8), 2850–2872. https://doi.org/10.3758/s13414-019-01792-7.
Lau, J.S.H., Brady, T.F., 2018. Ensemble statistics accessed through proxies: range heuristic and dependence on low-level properties in variability discrimination. J. Vis. 18 (9), 3. https://doi.org/10.1167/18.9.3, 3.
Lim, S.X.L., Höchenberger, R., Ruda, I., Fink, G.R., Viswanathan, S., Ohla, K., 2022. The capacity and organization of gustatory working memory. Sci. Rep. 12 (1), 8056. https://doi.org/10.1038/s41598-022-12005-x.
Luck, S.J., Vogel, E.K., 1997. The capacity of visual working memory for features and conjunctions. Nature 390, 279–281. https://doi.org/10.1038/36846.
Marchant, A.P., Simons, D.J., de Fockert, J.W., 2013. Ensemble representations: effects of set size and item heterogeneity on average size perception. Acta Psychol. 142 (2), 245–250. https://doi.org/10.1016/j.actpsy.2012.11.002.
Maule, J., Franklin, A., 2016. Accurate rapid averaging of multihue ensembles is due to a limited capacity subsampling mechanism. JOSA A 33 (3), A22–A29. https://doi.org/10.1364/JOSAA.33.000A22.
Morey, R.D., 2008. Confidence Intervals from Normalized Data: A correction to Cousineau (2005). Tutorial in Quantitative Methods for Psychology 4 (3), 61–64. https://doi.org/10.20982/tqmp.04.2.p061.
Papenmeier, F., Timm, J.D., 2021. Do group ensemble statistics bias visual working memory for individual items? A registered replication of Brady and Alvarez (2011). Atten. Percept. Psychophys. 83, 1329–1336. https://doi.org/10.3758/s13414-020-02209-6.
Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Katsman, E., Lindeløv, J.K., 2019. PsychoPy2: experiments in behavior made easy. Behav. Res. Methods 51, 195–203. https://doi.org/10.3758/s13428-018-01193-y.
Prolific. (2021). Oxford, UK. Available at: https://www.prolific.co.
Qualtrics. (2020). Provo, Utah, USA. Available at: https://www.qualtrics.com.
Spence, C., 2015. On the psychological impact of food colour. Flavour 4 (1), 1–16. https://doi.org/10.1186/s13411-015-0031-3.
Spence, C., Levitan, C.A., Shankar, M.U., Zampini, M., 2010. Does food color influence taste and flavor perception in humans? Chemosensory Perception 3 (1), 68–84. https://doi.org/10.1007/s12078-010-9067-z.
Utochkin, I.S., Brady, T.F., 2020. Independent storage of different features of real-world objects in long-term memory. J. Exp. Psychol. Gen. 149 (3), 530. https://doi.org/10.1037/xge0000664.
Valyear, K.F., Culham, J.C., Sharif, N., Westwood, D., Goodale, M.A., 2006. A double dissociation between sensitivity to changes in object identity and object orientation in the ventral and dorsal visual streams: a human fMRI study. Neuropsychologia 44 (2), 218–228. https://doi.org/10.1016/j.neuropsychologia.2005.05.004.
Wagenmakers, E., Wetzels, R., Borsboom, D., van der Maas, Han, L.J., 2011. Why psychologists must change the way they analyze their data: the case of psi: comment on bem (2011). J. Pers. Soc. Psychol. 100 (3), 426–432. https://doi.org/10.1037/a0022790.
Whitney, D., Yamanashi Yamanashi Leib, A., 2018. Ensemble perception. Annu. Rev. Psychol. 69, 105–129. https://doi.org/10.1146/annurev-psych-010416-044232.
Williams, R.S., Pratt, J., Ferber, S., Cant, J.S., 2021. Tuning the ensemble: incidental skewing of the perceptual average through memory-driven selection. J. Exp. Psychol. Hum. Percept. Perform. 47 (5), 648. https://doi.org/10.1037/xhp0000907.
Xu, Z., Adam, K.C.S., Fang, X., Vogel, E.K., 2018. The reliability and stability of visual working memory capacity. Behav. Res. Methods 50, 576–588. https://doi.org/10.3758/s13428-017-0886-6.
Yamanashi Leib, A.Y., Kosovicheva, A., Whitney, D., 2016. Fast ensemble representations for abstract visual impressions. Nat. Commun. 7 (1), 1–10. https://doi.org/10.1038/ncomms13186.
Yamanashi Leib, A., Chang, K., Xia, Y., Peng, A., Whitney, D., 2020. Fleeting impressions of economic value via summary statistical representations. J. Exp. Psychol. Gen. https://doi.org/10.1037/xge0000745.