



Published in final edited form as:

Psychol Sci. 2015 September ; 26(9): 1511–1521. doi:10.1177/0956797615592394.

Crowding in visual working memory reveals its spatial resolution and the nature of its representations

Benjamin J. Tamber-Rosenau, Anat R. Fintzi, and René Marois

Department of Psychology, Vanderbilt University, Nashville, TN

Abstract

Spatial resolution fundamentally limits any image representation. While this limit has been extensively investigated for perceptual representations by assessing how neighboring flankers degrade the perception of a peripheral target with visual crowding, the corresponding limit for representations held in visual working memory (VWM) is unknown. Here we evoked crowding in VWM and directly compared its resolution to that of perception. Remarkably, the spatial resolution of VWM proved no worse than that of perception. However, mixture modeling of errors due to crowding revealed the qualitatively distinct nature of these representations. Perceptual crowding errors arose from both increased imprecision in target representations and substitution of flankers for targets. By contrast, VWM crowding errors exclusively arose from substitutions, suggesting that VWM transforms analog perceptual representations into discrete items. Thus, while perception and VWM share a common resolution limit, exceeding this limit reveals distinct mechanisms for perceiving images and holding them in mind.

Our perception of the visual world is limited by our ability to resolve its elements. This is easily demonstrated by comparing scene perception across visual eccentricities: while we easily individuate and identify foveated items, the same image becomes blurry and amorphous in the periphery.

Though limits on the spatial resolution of perceptual representations have been extensively studied (e.g., Anton-Erxleben & Carrasco, 2013; Whitney & Levi, 2011), this is not so for representations maintained in visual working memory (VWM) after sensory input has faded. A decade of research has revealed that *object features* are degraded in VWM relative to perception (Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Fougnie, Asplund, & Marois, 2010; Fougnie, Suchow, & Alvarez, 2012; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004; Zhang & Luck, 2008), but it is unknown if the *spatial resolution* of VWM is comparably degraded. Ben-Shalom and Ganel (2014) recently measured the precision of VWM distance representations but not the spatial resolution of VWM, leaving unanswered whether spatial proximity differentially impairs our ability to resolve items in VWM and perception.

Corresponding authors: Benjamin J. Tamber-Rosena, benjamin.j.tamber-rosenau@vanderbilt.edu, 627 Wilson Hall, Vanderbilt University Department of Psychology, PMB 407817, 2301 Vanderbilt Place, Nashville, TN 37240-7817, René Marois, rene.marois@vanderbilt.edu, 530 Wilson Hall, Vanderbilt University Department of Psychology, PMB 407817, 2301 Vanderbilt Place, Nashville, TN 37240-7817.

A well-known means to assess the spatial resolution of perception (Whitney & Levi, 2011) and attention (He, Cavanagh, & Intriligator, 1996) is the visual crowding paradigm. In crowding, perceptual representations of targets presented in the periphery are degraded by flanking items (Bouma, 1970; Levi, 2008; Whitney & Levi, 2011). Critically, the target-flanker distance regulates the degree of interference, revealing the limit of perceptual spatial resolution (Bouma, 1970; Levi, 2008; Levi, Hariharan, & Klein, 2002). As such, crowding represents a potentially excellent means for comparing the spatial resolution of VWM to that of perception. Moreover, studying how crowding degrades items can reveal much about the nature of VWM representations, just as it has done for perceptual representations. For visual perception, crowding is thought to degrade image representation in one or both of two ways (Levi, 2008; Whitney & Levi, 2011). First, target features may be averaged with or otherwise contaminated by flanker features (cross-item pooling error), leading to greater imprecision. Second, targets and flankers may be correctly individuated while lacking positional fidelity, resulting in a flanker being confused for a target at report (substitution error). These two types of errors can be distinguished using mixture modeling, a technique that discerns the relative contributions of multiple sources of information and error to the overall response distribution. Indeed, recent studies suggest that both pooling and substitution errors underlie crowding in perception (Ester, Klee, & Awh, 2014; Freeman, Chakravarthi, & Pelli, 2012).

The goal of the present study was to evoke crowding in VWM in order to characterize its spatial resolution and compare the effects of VWM crowding to perceptual crowding. We adapted a standard perceptual crowding paradigm to VWM and measured how target-report errors changed with target-flanker distance. Strikingly, we found that the spatial resolution limit of VWM was no worse than that of perception. However, mixture-modeling analyses (Bays et al., 2009; Zhang & Luck, 2008) of the consequences of exceeding such limits revealed the qualitatively distinct natures of perceptual and VWM representations.

Method

Subjects

Twelve subjects completed Experiment 1 and six subjects completed Experiment 2. In Experiment 1, an additional three subjects were terminated prior to collection of a full data set due to failure to fixate consistently. In Experiment 2, an additional two subjects were rejected without early termination, also due to failures to fixate consistently. No subject participated in both experiments. All subjects gave written informed consent as approved by the Vanderbilt University Institutional Review Board. Subjects were paid \$12/hour for participation.

Eyetracking

We monitored eye position using an Arrington PC-60 eyetracker controlled by Viewpoint software, the Viewpoint Matlab toolbox, and custom Matlab code. Trials in which we detected eye movements were rejected from all analyses. Detailed eyetracking methods and analyses are included in the Supplemental Material.

General Task Design and Procedure

The basic task design consisted of a standard crowding paradigm in which subjects had to report a feature of one of three simple oriented bars presented in the display (Figure 1). Across trials, we factorially varied the report feature (orientation versus location), representation level (perceptual versus VWM), and amount of crowding, i.e., inter-item distance (throughout this paper we refer to this inter-item distance manipulation as a manipulation of crowding, and the effects of crowding can be quantified as differences in task performance across levels of inter-item distance). In addition, and unlike most crowding paradigms in which only the central item is ever reported, we also varied which of the three bars served as a target on any given trial in order to force subjects to maintain all item locations during VWM trials. This four-factor task design meant that we could only acquire few trials per cell in each experimental session. To obtain sufficient trials for modeling, we therefore required subjects to perform numerous sessions, as detailed further below.

On each trial, subjects viewed a central fixation dot, a stimulus array, a visual cue, and an adjustment item; they also heard an auditory cue. All stimuli were presented on a black background. The stimulus array consisted of three oriented white bars (length: 1.69 degrees of visual angle, DVA; width: 0.21 DVA) with equal spacing along an imaginary horizontal line 12.20 DVA above a central white fixation dot. Bar orientations varied pseudo-randomly across items in increments of 10 degrees of rotational angle (DRA) in the range ± 45 DRA from vertical, with the constraint that no two stimuli on the same trial had the identical orientation. We chose to use this restricted range so that we could maximize the effects of crowding by presenting items as close as possible without touching each other. The horizontal center of the stimulus array was also varied randomly across trials.

On perceptual trials (Fig. 1 panels a and c), the stimulus array was displayed until response. A 500 ms auditory cue indicating the feature to be reported (orientation versus location; the spoken words “tilt,” or “place,” respectively) began after 1 s of stimulus display. Immediately following the offset of the auditory cue, we presented a visual cue that indicated which one of the three bars to report. At the same time, we presented an adjustment item that was used for target report (see below). The visual cue and adjustment item remained on screen until the subject finalized his or her response.

On VWM trials, the stimulus array offset after 1 s. That offset was followed by an 800 ms blank delay before the onset of the 500 ms auditory cue that signaled which feature to report. The offset of the auditory cue was followed immediately by the onset of the visual cue that indicated which one of the three bars to report. Simultaneously with the visual cue, the adjustment item that was used for target report (see below) appeared. Thus, the total VWM delay was 1300 ms. As for perceptual trials, the visual cue and adjustment item remained on screen until the subject finalized his or her response.

For both the perceptual and VWM conditions, all target item positions and features (orientation versus location) were cued with equal frequency. This paradigm enforced the use of VWM representations that maintained the original, crowded perceptual conditions during the delay interval because subjects did not know which of the three items and which

of the two features—orientation or location—they would be asked to report until after the delay period.

Following their response, subjects in both experiments (save for the first three in Experiment 1) were presented with a 500 ms long feedback screen reporting their error in DRA (orientation trials) or pixels (location trials). Subjects were instructed to do their best to minimize errors, using the feedback information.

During the first session of each experiment, subjects performed at least three practice runs. These simplified and shortened runs gradually introduced various features of the task. Subjects were instructed to perform additional practice runs until they and the experimenter were confident in their understanding of all cues, stimuli, and trial types. Subjects completed a mean of 4.1 (SD: 0.5) practice runs in Experiment 1, and 3.3 (SD: 0.5) practice runs in Experiment 2. All practice run data were discarded and not analyzed. All task displays and response acquisitions were accomplished via custom Matlab code using the Psychophysics Toolbox version 3 (Kleiner, Brainard, & Pelli, 2007).

Visual cue and adjustment item: Orientation-report trials—Orientation trials used a peripheral red dot as the visual cue, and a centrally presented bar (similar to the target and flanker bars) as the adjustment item. The identity of the target bar on orientation trials was signaled by the location of the red dot, which was displayed 1.83 DVA directly below the target item. The red dot visual cue thus indicated the spatial location of the target bar but was uninformative as to the orientation of the target bar.

Subjects reported the orientation of the target bar by adjusting the orientation of the central adjustment bar displayed at fixation. The starting orientation of this central bar adjustment item was randomly chosen from the range ± 55 DRA in increments of 10 DRA. To adjust its orientation, subjects repeatedly pressed the “<” or “>” keys, with each press tilting the central bar 10 DRAs. When subjects arrived at a satisfactory answer, i.e., had matched the central adjustment bar orientation to the perceived or memorized target orientation, they committed their response by pressing the quotation-mark key.

Visual cue and adjustment item: Location-report trials—The identities of the visual cue and adjustment item switched on location trials. Specifically, the identity of the target was signaled by the orientation of a central bar cue, and subjects reported the target location by adjusting the position of a red dot adjustment item.

The central bar visual cue had the exact same orientation as the target bar but differed in that it was located at fixation. Thus, the central bar visual cue only indicated the target bar’s orientation, not its location. Subjects reported the location of the target bar by adjusting the horizontal position of the red dot adjustment item. The appearance of the red dot adjustment item was identical to that of the red dot on orientation-report items except that on location-report trials, the starting horizontal position of the red dot was chosen randomly (see below). To adjust the horizontal position of the red dot adjustment item, subjects repeatedly pressed the “<” or “>” keys, with each press moving the dot 0.17 DVA. When subjects arrived at a satisfactory answer, they committed their response by pressing the quotation-mark key.

Experiment 1

The design of Experiment 1 was as described in the General Task Design section with the following additional characteristics: The horizontal center of the stimulus array was placed randomly in the range ± 5.43 DVA from the horizontal center of the screen (i.e. fixation point). On location-report trials, the red dot adjustment item's starting location was random within the range ± 5.97 DVA from the horizontal center of the screen.

Experiment 1 was a 2 (representation level) x 2 (crowding distance, either 1.36 or 5.43 DVA) x 2 (report feature) x 3 (target item) design with a total of 24 cells. An average hour-long task session in Experiment 1 contained approximately 12 trials per cell, prior to excluding trials for breaks in fixation. We thus required subjects to attend a series of sessions in order to obtain sufficient trials in each cell for modeling. Subjects performed a mean of 12.1 (SD: 1.4) sessions, each containing 3.6 (SD: 0.35) task runs of 80 trials per run. After rejecting trials on which fixation was broken, we obtained an average of 110.21 (SD: 20.76) trials per cell for the critical cells of the design (central-target orientation trials). Further details on trial counts, including trial rejection rates due to fixation breaks, may be found in Table S1.

Experiment 2

The purpose of Experiment 2 was to ensure that the results of Experiment 1 would generalize to other crowding distances. Thus, Experiment 2 employed three levels of crowding distance, 1.36, 3.39, or 7.46 DVA. To accommodate the larger range of crowding distances, we expanded the range in which the horizontal center of the stimulus array could be placed. Specifically, the horizontal center of the stimulus array was placed randomly in the range ± 7.46 DVA from the horizontal center of the display. On location-report trials, the red dot adjustment item's starting location was completely random within the range ± 8.20 DVA from the horizontal center of the screen. All other stimulus characteristics were identical to Experiment 1.

Experiment 2 was a 2 (representation level) x 3 (crowding distance) x 2 (report feature) x 3 (target item) design with a total of 36 cells. An average hour-long task session in Experiment 2 contained approximately 8 trials per cell, prior to excluding trials for breaks in fixation. We thus required subjects to attend a number of sessions in order to obtain sufficient trials in each cell for modeling. Subjects performed a mean of 24.3 (SD: 4.4) sessions of 2.8 (SD: 0.15) task runs with each run containing 96 trials. After rejecting trials on which fixation was broken, we obtained an average of 128.61 (SD: 23.09) trials per cell for the critical cells of the design (central-target orientation trials). Further details on trial counts, including trial rejection rates due to fixation breaks, may be found in Table S2.

Analysis of report errors

All analyses were conducted separately for each experiment using custom code implemented in Matlab. We only analyzed trials in which the central item was probed because central targets exhibit the greatest visual crowding effects (Levi, 2008). Though our design included location-report trials in order to force subjects to maintain the stimulus array in VWM in its original location, these trials are not analyzed because errors on location-report trials could

be due to either location errors or crowding of the orientation cue that indicated the target item.

First, we ran an ANOVA on the non-directional report error magnitudes broken down by representation level and crowding distance. We then fit directional target orientation report errors with various mixture models (Bays et al., 2009; van den Berg et al., 2012; Zhang & Luck, 2008; see below). Models were modified from their original forms to remove parameters that varied with set size, as our stimulus arrays always contained three items (one target and two flankers). In addition, each model was adapted to use truncated normal distributions rather than circular (von Mises) distributions for target precision. We chose to use truncated normal distributions because, unlike most VWM mixture modeling studies, our paradigm required subjects to report feature values over a restricted range that did not “wrap” in a circle. Such a restricted orientation range was necessary to maximize item proximity, and thus, crowding. Unlike the circular normal (von Mises) probability distributions used in most implementations of these models, a truncated normal distribution is bounded. Preliminary examination of our data revealed that the most extreme response on any trial from any subject had an absolute value of 65 DRA relative to vertical. Thus, we bounded our truncated normal distribution at ± 75 DRA relative to vertical (modeling using truncation at ± 89 DRA yielded similar results). This choice ensured that no data were excluded from analysis while avoiding using such a wide distribution as to deflate the possibility of obtaining a non-zero guess rate parameter estimate in those models that included a guess rate.

We fit adaptations of the model of Zhang and Luck (2008) and two variants of the model of Bays et al. (2009). For all models, we fixed the representation distribution mean parameter at the veridical orientation value of the target stimulus. Our implementation of the Zhang and Luck (2008) model included an imprecision parameter (i.e., the standard deviation of the modeled truncated normal distribution of the target representation) and a guess rate (proportion of reports drawn from a uniform distribution) parameter. The guess rate corresponds to a complete failure to represent an item, resulting in a random guess as to its orientation. We also fit two variants of the Bays et al. (2009) model. The first, “no guess,” variant included an imprecision parameter (i.e., the standard deviation of the modeled truncated normal distribution of the target orientation representation) and a substitution rate parameter (i.e., the proportion of trials on which report was drawn from a flanker’s representational distribution rather than that of the target). Unlike in the Zhang and Luck (2008) model, the “no guess” Bays et al. (2009) model also fits a separate distribution for each flanker representation. The target and flanker representations are constrained to share a common standard deviation, i.e., imprecision parameter, but they are centered on the veridical target and flanker feature values, respectively. In a second, “combined,” variant of the Bays model, we added a guess rate (uniform distribution) parameter as in Zhang and Luck (2008). This model was otherwise identical to the “no guess” Bays model. We also considered additional model variants – namely the variable precision model of van den Berg et al. (2012) (see also Fougne et al., 2012) – but opted not to use them because their multi-component variable-precision parameters do not clearly map onto an interpretable cognitive construct in the way that a fixed imprecision itself does, or in the way that substitution rate (in the Bays model; a confusion between two represented items) or guess rate (in the Zhang

and Luck model; a failure to encode or maintain an item) do. Moreover, a preliminary application of the variable precision model did not reveal any benefit to using this model compared to others we tested.

Model selection—Although previous modeling of visual crowding (Ester et al., 2014) favored the “combined” Bays model that included imprecision and substitution parameters, the “no guess” Bays model produced the most plausible and internally consistent parameter estimates on our data. Specifically, in models including a guess parameter, guess rates and imprecision traded off idiosyncratically across subjects in many experimental conditions (see Supplemental Figure S2–Figure S4; also see Supplement text and Fig. S5–Fig. S6 for further evidence that guessing does not drive the present results). Such parameter trade-off is a hallmark of overfitting (Pitt & Myung, 2002), i.e., fitting subject-specific noise variance rather than arriving at useful parameter estimates of the true signal. In addition to subject-specific overfitting, systematic inconsistencies emerged from the application of the Zhang & Luck and “combined” Bays models to our data. In particular, the parameter estimates suggested that VWM representational fidelity increased, though with implausibly high guess rates, for more crowded conditions. These results are neither predicted nor realistic under any account of VWM or crowding of which we are aware. A probable reason why guesses appear to play a much-reduced role in our data compared to Ester et al. (2014) lies in the stimulus presentation methodology: Ester et al. (2014) presented their stimuli for a 75 ms encoding period, likely leading to frequent trials in which some stimuli failed to be encoded at all. By contrast, we provided a minimum of 1 s of encoding time, thus reducing the likelihood of a total failure to encode any item (Bays et al., 2009). These extended viewing times do not abolish crowding (Intriligator & Cavanagh, 2001; Townsend, Taylor, & Brown, 1971).

Given that the “no guess” Bays model provided the most plausible fit to the data, we used its parameter estimates for further statistical analyses. Specifically, we extracted subject- and condition- (representation level x crowding distance) specific parameter estimates from this model and subjected them to separate ANOVAs in which we treated subject as a random effect.

Results

Non-directional report error

To assess whether crowding affected perceptual and VWM representations differently, we first considered non-directional report error magnitudes (Figure 2; see also Supplemental Figure S1) using ANOVAs with crowding distance and representation level (perceptual versus VWM) as factors. ANOVA results are presented in Table 1 (Expt. 1) and Table 2 (Expt. 2).

In both experiments, the ANOVAs revealed a main effect of crowding – such that error magnitudes increased with shorter inter-item distance, consistent with crowding predictions – and a main effect of representation level – such that errors were larger under VWM than under perception. Importantly, Experiment 1 showed no evidence of an interaction: the amount of crowding in VWM was indistinguishable from that in perception. This interaction

is the critical test for differential spatial resolution in perception and VWM because the interaction measures whether identical changes in inter-stimulus distance lead to differential crowding effects in perception and VWM. In Experiment 2, the interaction achieved statistical significance, but appears to have been driven by a floor effect on error in the low crowding (high inter-item distance) perceptual condition. This conjecture is bolstered by the absence of an interaction in a separate ANOVA that only considered the medium and high crowding conditions (interaction: $F(1,5)=1.24$, $p=0.3157$, $\eta^2_p=0.1588$; main effect of crowding: $F(1,5)=149.28$, $p=6.4950\times 10^{-5}$, $\eta^2_p=0.7404$; main effect of representation: $F(1,5)=82.49$, $p=0.0003$, $\eta^2_p=0.9053$). The floor effect for the low crowding perceptual condition is unsurprising given that the stimulus separation was 7.46 DVA, which translates to 0.61 times the stimulus eccentricity. Since the critical distance for experiencing crowding in visual perception is typically reported as between 0.1 and 0.5 times the stimulus eccentricity (Bouma, 1970; Levi et al., 2002), error in this condition of Expt. 2 should primarily be driven by the limits of featural precision in peripheral vision for single items that have been documented previously (e.g., Ester et al., 2014). Thus, the floor effect for the high inter-item distance (low crowding) perceptual condition of Expt. 2 leads to an underestimation of the size of the perceptual crowding effect for high versus intermediate inter-item distance trials in Expt. 2.

We should also note that the 10 DRA of non-directional error in Experiment 2's low crowding VWM condition should not be taken as evidence that VWM representations are strongly crowded in this condition. Instead, non-directional error here primarily reflects the level of imprecision with which orientation is represented in VWM under conditions of minimal crowding, comparable to measurement of the representational precision of a VWM target in isolation. Indeed, our VWM low crowding conditions yielded roughly comparable measures of orientation representational fidelity to those previously obtained in VWM for isolated targets or targets with a single distractor (Bays & Husain, 2008; Wilken & Ma, 2004). Put differently, the error in each individual condition tells us about the fidelity of the *orientation feature representation* in that condition, while the difference in error between conditions with different target-flanker spacing tells us about the influence of *crowding*, i.e., *spatial resolution*, on that feature representation.

Summing up the non-directional error results, across the two experiments we can safely conclude that inter-item distance and representation level did not meaningfully interact. In other words, perception and VWM share a common spatial resolution limit, at least with the granularity of the target-flanker spacings we have tested.

Representational imprecision and substitution

While the results of the non-directional error analysis suggested that manipulations of inter-item distance have the same crowding effect on both perception and VWM, unpacking target errors into imprecision (pooling) and substitution with the “no guess” Bays mixture model revealed that crowding impacts perception and VWM in qualitatively distinct ways (Figure 3; also see Supplemental Figure 5). Specifically, separate ANOVAs on imprecision and substitution errors (Expt. 1: Table 1; Expt. 2: Table 2) not only revealed main effects in both experiments, but also interactions between crowding distance and representation level

for both imprecision and substitution errors. The pattern of these interactions (Figure 3) was such that perceptual crowding increased both imprecision and substitution, whereas VWM crowding only increased substitution, leaving the precision of features intact. One possible account of these results is that location representations might be less precise in VWM than in perception, leading subjects to confuse which item was cued on VWM trials and thus causing apparent substitution errors. Were this the case, non-directional report errors would be expected to show greater crowding effects in VWM than in perception. Instead, we observed comparable crowding in both representation levels, inconsistent with this trivial explanation of the modeling results.

To better understand the interactions obtained from the mixture modeling, we next performed tests of simple main effects to separately assess the consequences of crowding in VWM and perception. We used paired t-tests to assess Experiment 1, which had only two levels of crowding distance, and one-way ANOVAs to assess Experiment 2, which had three levels of crowding. We first assessed the simple main effect of crowding distance in VWM on imprecision parameter estimates. We did not observe any significant effect of crowding distance on imprecision in VWM in either experiment (Expt. 1: $t(11)=1.1901$, $p=0.2591$, Cohen's $d=0.3436$; Expt. 2: $F(2,10)=1.2037$, $p=0.3401$, $\eta^2_p=0.0241$). However, we observed large increases in imprecision with decreased perceptual crowding for both experiments (Expt. 1: $t(11)=7.1836$, $p=1.7898\times 10^{-5}$, Cohen's $d=2.0737$; Expt. 2: $F(2,10)=30.0764$, $p=5.8854\times 10^{-5}$, $\eta^2_p=0.9572$). These results support our conclusion from the main ANOVAs that, unlike perceptual crowding, VWM crowding did not modulate representational precision, i.e., did not lead to pooling of target and flanker feature values.

We next performed parallel tests on substitution parameter estimates. We observed significant effects of VWM crowding distance on substitution in both experiments (Expt. 1: $t(11)=15.5697$, $p=7.693\times 10^{-9}$, Cohen's $d=4.4946$; Expt. 2: $F(2,10)=136.9534$, $p=5.4215\times 10^{-8}$, $\eta^2_p=0.7467$). We also observed significant effects of perceptual crowding distance on substitution in both experiments (Expt. 1: $t(11)=5.0561$, $p=3.6853\times 10^{-4}$, Cohen's $d=1.4596$; Expt. 2: $F(2,10)=12.8569$, $p=0.0017$, $\eta^2_p=0.8486$). These results support our conclusion from the main ANOVAs that crowding modulated substitution in both VWM and perception, though the interactions in the main ANOVAs also indicate that crowding had a greater effect on substitutions in VWM than in perception.

Discussion

Here we evoke, for the first time, VWM crowding and show that, contrary to expectations, the spatial resolution of VWM is no worse than that of perception. However, mixture modeling of report errors indicated that exceeding spatial resolution limits degrades perceptual and mental representations in qualitatively different ways.

That VWM is subject to similar spatial resolution limits as perception accords with the *sensory recruitment hypothesis* that VWM representations are perceptual representations maintained after stimulus offset (Ester, Serences, & Awh, 2009; Serences, Ester, Vogel, & Awh, 2009; also see Tsubomi, Fukuda, Watanabe, & Vogel, 2013). However, were VWM simply time-extended perception, errors from perception and VWM would not be

categorically distinct. Instead, we show that identical crowding leads to dissociable errors for perception and VWM – both imprecision and substitution in perception, but exclusively substitution in VWM. Thus, contrary to a strong form of the sensory recruitment hypothesis, our results indicate that VWM representations may be significantly transformed from perceptual representations.

If substitution errors reflect report of a non-target item, increased substitution errors resulting from exceeding the limits of spatial resolution in VWM could be taken as evidence in support of a “slot” model of VWM (e.g., Luck & Vogel, 1997). Slot models posit that VWM items are represented using one of a few discrete, indivisible units of resource. It is also possible, however, that substitutions reflect feature-binding errors such that the location of the target is erroneously bound to the orientation of a flanker (Levi, 2008; Pelli, Palomares, & Majaj, 2004), akin to the perceptual phenomenon of illusory conjunctions (Treisman & Schmidt, 1982; Wheeler & Treisman, 2002). Further research is necessary to adjudicate between these alternatives.

What seems certain is that exceeding spatial resolution limits in VWM leads exclusively to substitutions of intact features, while exceeding spatial resolution limits in perception also leads to pooling of feature values across items. How can crowding have such different effects on perception and VWM? We propose that VWM transforms continuous analog perceptual representations into discrete digital mental representations, with these discrete VWM representations suffering exclusively from all-or-none feature or object substitution under crowded conditions. We further suggest that items that are less precisely represented in perception are more susceptible to substitution, accounting for the shift in error type with the transition from perception to VWM (see also Brady, Konkle, Gill, Oliva, & Alvarez, 2013). Evidently, while perception and VWM share the same spatial resolution, the limit of this resolution reveals distinct mechanisms by which we perceive images and hold their representations in mind. It will be up to neurobiological inquiries to reveal the nuts and bolts of these perceptual and VWM representations (e.g., Ester et al., 2013; Sprague et al., 2014).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Acknowledgements and Funding

We thank Shea Littlepage and Mareike Eydt-Beebe for technical assistance. We used Vanderbilt Advanced Computing Center for Research and Education resources and support from NSF Graduate Research Fellowship DGE1445197 to A.F. and NIH Core Grant P30EY008126 to the Vanderbilt Vision Research Center.

References

- Anton-Erxleben K, Carrasco M. Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. *Nat Rev Neurosci.* 2013; 14(3):188–200. [PubMed: 23422910]
- Bays PM, Catalao RF, Husain M. The precision of visual working memory is set by allocation of a shared resource. *J Vis.* 2009; 9(10):1–11. [PubMed: 19810788]

- Bays PM, Husain M. Dynamic shifts of limited working memory resources in human vision. *Science*. 2008; 321(5890):851–854. [PubMed: 18687968]
- Ben-Shalom A, Ganel T. Spatial resolution in visual memory. *Psychon Bull Rev*. 2014
- Bouma H. Interaction effects in parafoveal letter recognition. *Nature*. 1970; 226(5241):177–178. [PubMed: 5437004]
- Brady TF, Konkle T, Gill J, Oliva A, Alvarez GA. Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*. 2013; 24(6):981–990. [PubMed: 23630219]
- Ester EF, Anderson DE, Serences JT, Awh E. A neural measure of precision in visual working memory. *J Cogn Neurosci*. 2013; 25(5):754–761. [PubMed: 23469889]
- Ester EF, Klee D, Awh E. Visual crowding cannot be wholly explained by feature pooling. *J Exp Psychol Hum Percept Perform*. 2014; 40(3):1022–1033. [PubMed: 24364703]
- Ester EF, Serences JT, Awh E. Spatially Global Representations in Human Primary Visual Cortex during Working Memory Maintenance. *Journal of Neuroscience*. 2009; 29(48):15258–15265. [PubMed: 19955378]
- Fougnie D, Asplund CL, Marois R. What are the units of storage in visual working memory? *Journal of Vision*. 2010; 10(12)
- Fougnie D, Suchow JW, Alvarez GA. Variability in the quality of visual working memory. *Nat Commun*. 2012; 3:1229. [PubMed: 23187629]
- Freeman J, Chakravarthi R, Pelli DG. Substitution and pooling in crowding. *Atten Percept Psychophys*. 2012; 74(2):379–396. [PubMed: 22160819]
- He S, Cavanagh P, Intriligator J. Attentional resolution and the locus of visual awareness. *Nature*. 1996; 383(6598):334–337. [PubMed: 8848045]
- Intriligator J, Cavanagh P. The spatial resolution of visual attention. *Cogn Psychol*. 2001; 43(3):171–216. [PubMed: 11689021]
- Kleiner M, Brainard D, Pelli D. What's new in Psychtoolbox-3? *Perception*. 2007; 36:14–14.
- Levi DM. Crowding--an essential bottleneck for object recognition: a mini-review. *Vision Res*. 2008; 48(5):635–654. [PubMed: 18226828]
- Levi DM, Hariharan S, Klein SA. Suppressive and facilitatory spatial interactions in peripheral vision: peripheral crowding is neither size invariant nor simple contrast masking. *J Vis*. 2002; 2(2):167–177. [PubMed: 12678590]
- Luck SJ, Vogel EK. The capacity of visual working memory for features and conjunctions. *Nature*. 1997; 390(6657):279–281. [PubMed: 9384378]
- Pelli DG, Palomares M, Majaj NJ. Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J Vis*. 2004; 4(12):1136–1169. [PubMed: 15669917]
- Pitt MA, Myung IJ. When a good fit can be bad. *Trends Cogn Sci*. 2002; 6(10):421–425. [PubMed: 12413575]
- Serences JT, Ester EF, Vogel EK, Awh E. Stimulus-Specific Delay Activity in Human Primary Visual Cortex. *Psychological Science*. 2009; 20(2):207–214. [PubMed: 19170936]
- Sprague TC, Ester EF, Serences JT. Reconstructions of information in visual spatial working memory degrade with memory load. *Curr Biol*. 2014; 24(18):2174–2180. [PubMed: 25201683]
- Townsend JT, Taylor SG, Brown DR. Lateral Masking for Letters with Unlimited Viewing Time. *Perception & Psychophysics*. 1971; 10(5):375.
- Treisman A, Schmidt H. Illusory Conjunctions in the Perception of Objects. *Cognitive Psychology*. 1982; 14(1):107–141. [PubMed: 7053925]
- Tsubomi H, Fukuda K, Watanabe K, Vogel EK. Neural Limits to Representing Objects Still within View. *J Neurosci*. 2013; 33(19):8257–8263. [PubMed: 23658165]
- van den Berg R, Shin H, Chou WC, George R, Ma WJ. Variability in encoding precision accounts for visual short-term memory limitations. *Proc Natl Acad Sci U S A*. 2012; 109(22):8780–8785. [PubMed: 22582168]
- Wheeler ME, Treisman AM. Binding in short-term visual memory. *Journal of Experimental Psychology-General*. 2002; 131(1):48–64. [PubMed: 11900102]

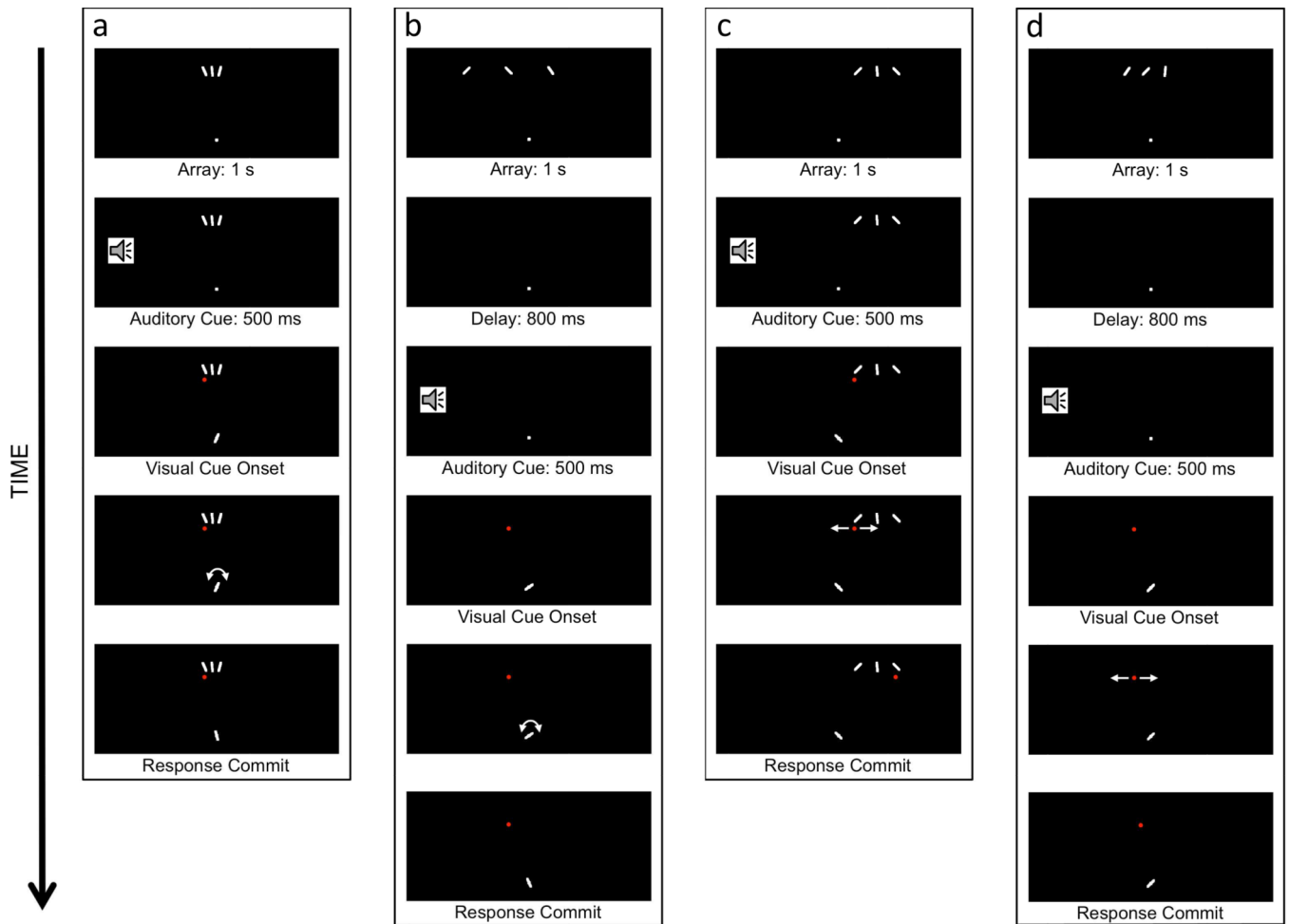
- Whitney D, Levi DM. Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends Cogn Sci.* 2011; 15(4):160–168. [PubMed: 21420894]
- Wilken P, Ma WJ. A detection theory account of change detection. *J Vis.* 2004; 4(12):1120–1135. [PubMed: 15669916]
- Zhang W, Luck SJ. Discrete fixed-resolution representations in visual working memory. *Nature.* 2008; 453(7192):233–235. [PubMed: 18385672]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Fig. 1.**

Task sequence examples. Task sequence proceeds from top to bottom in each panel. For all frames, the empty bottom portion of the display has been cropped. Note that the speaker icon and the white arrows indicating response adjustment did not appear in the paradigm and are included here for illustration only. All items are to scale except that the stimulus lines and fixation dot have been enlarged for visibility. **(a)** An orientation-report perceptual trial (shown with high crowding). A stimulus array of three oriented bars was presented and remained on screen until the end of a trial. After 1 sec, a 500 ms auditory cue instructed the subject to report the orientation of the target bar. After the auditory cue, a *visual cue* (a peripheral red dot) and an *adjustment item* (a centrally presented oriented bar) appeared. On such orientation-report trials, the red dot was positioned directly below the target, and subjects reported the target orientation by rotating the central bar adjustment item until its orientation matched that of the target. **(b)** A low crowding (Expt. 2) orientation-report VWM trial. Task sequence is identical to panel (a) except that the stimulus array offsets after 1 s. A further 800 ms delay period preceded the auditory cue. **(c)** A medium crowding (Expt. 2) location-report perceptual trial. Task sequence is identical to panel (a). Critically, in a reversal of their roles in the orientation-report trials, in location-report trials the peripheral red dot served as the adjustment item and the central oriented bar served as the visual cue.

Subjects horizontally translated the peripheral red dot adjustment item until it was directly beneath the target signaled by the central bar visual cue. **(d)** A medium crowding (Expt. 2) location-report VWM trial. Task sequence is as in (b).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Experiment 1

Experiment 2

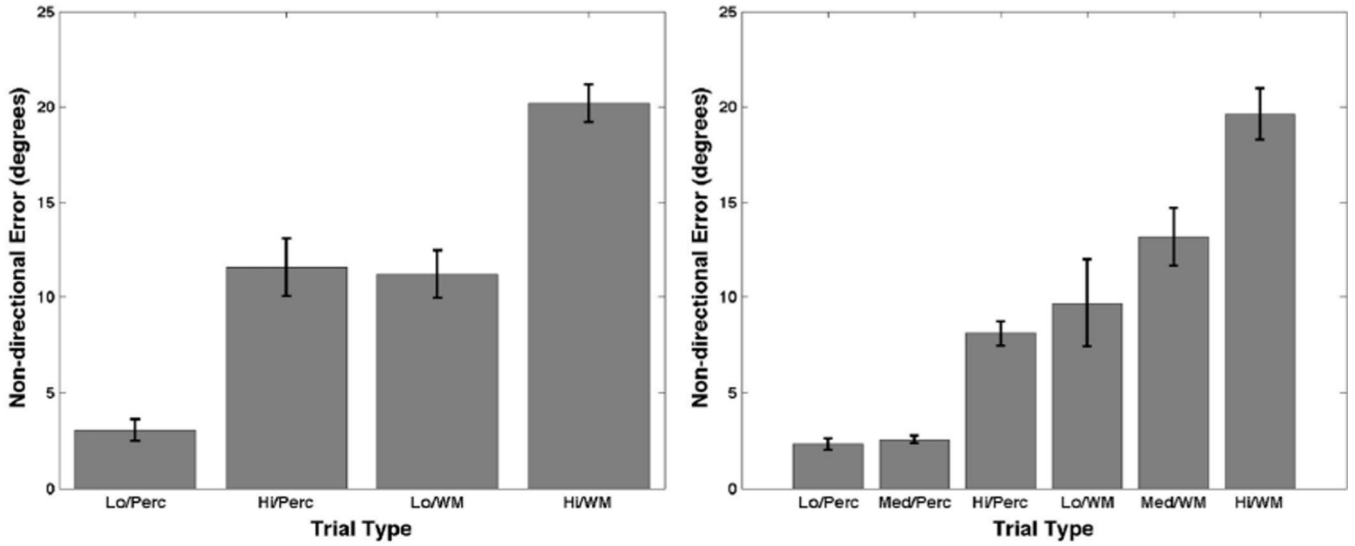


Fig. 2. Non-directional error in target report. In both experiments, errors increased with both crowding and dependence on working memory representations. Error bars represent standard error of the mean. **Left:** Expt. 1. **Right:** Expt. 2. **Legend:** “Lo” = low crowding; “Med” = medium crowding; “Hi” = high crowding; “Perc” = perceptual; “WM” = visual working memory.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

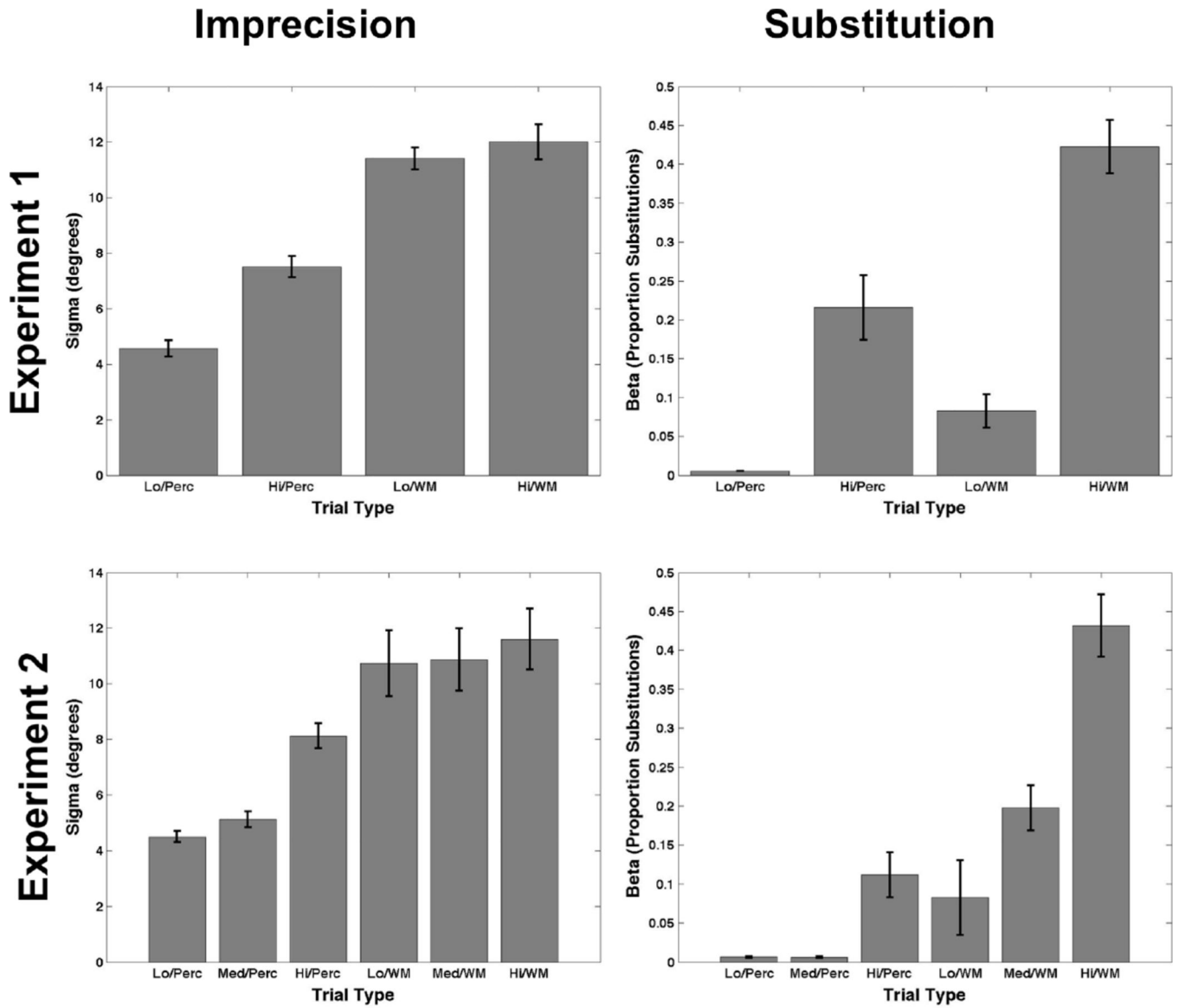


Fig. 3. Model parameter estimates from the “no guess” Bays model. In both experiments, crowding increased both imprecision and substitution in perception. However, crowding increased only substitution in working memory. Error bars represent standard error of the mean. **Top Row:** Expt. 1. **Bottom Row:** Expt. 2. **Left Column:** Imprecision (standard deviation of the truncated normal distribution). **Right Column:** Substitution rate (proportion of trials on which a flanker feature was reported instead of the target’s feature). Legend is as in Figure 2.

ANOVAs on Experiment 1 Errors and Parameter Estimates. d.f., degrees of freedom; Num., numerator; Denom., denominator; η^2_p , partial eta squared.

Table 1

Factor	F-ratio	d.f. Num.	d.f. Denom.	p	η^2_p
Non-directional report error					
Crowding Distance	155.16	1	11	7.9202×10^{-8}	0.7887
Representation Level	191.00	1	11	2.6908×10^{-8}	0.7845
Crowding Dist. x Repr. Level	1.83	1	11	0.2029	0.1015
Imprecision					
Crowding Distance	27.93	1	11	0.0003	0.4763
Representation Level	128.43	1	11	2.0898×10^{-7}	0.9033
Crowding Dist. x Repr. Level	14.33	1	11	0.0030	0.5285
Substitution					
Crowding Distance	107.83	1	11	5.0670×10^{-7}	0.7811
Representation Level	57.89	1	11	1.0490×10^{-5}	0.4882
Crowding Dist. x Repr. Level	10.39	1	11	0.0081	0.3513

Table 2

ANOVAs on Experiment 2 Errors and Parameter Estimates. d.f., degrees of freedom; Num., numerator; Denom., denominator; η^2_p , partial eta squared.

Factor	F-ratio	d.f. Num.	d.f. Denom.	p	η^2_p
Non-directional report error					
Crowding Distance	147.75	2	10	3.7575×10 ⁻⁸	0.7334
Representation Level	43.57	1	5	0.0012	0.8531
Crowding Dist. x Repr. Level	7.25	2	10	0.0113	0.6497
Imprecision					
Crowding Distance	24.21	2	10	0.0001	0.3526
Representation Level	27.28	1	5	0.0034	0.7892
Crowding Dist. x Repr. Level	5.84	2	10	0.0209	0.6450
Substitution					
Crowding Distance	108.48	2	10	1.6606×10 ⁻⁷	0.8184
Representation Level	27.69	1	5	0.0033	0.8228
Crowding Dist. x Repr. Level	28.17	2	10	7.7822×10 ⁻⁵	0.8520