



# A direct comparison of central tendency recall and temporal integration in the successive field iconic memory task

Jacob Zepp<sup>1</sup> · Chad Dubé<sup>1</sup> · David Melcher<sup>2,3</sup>

Accepted: 21 October 2020 / Published online: 3 January 2021  
© The Psychonomic Society, Inc. 2021

## Abstract

The ensemble coding literature suggests the existence of a fast, automatic formation of some ensemble codes. Can statistical representations, such as memory for the central tendency along a particular visual feature dimension, be extracted from information held in the sensory register? Furthermore, can knowledge of early, iconic memory processes be used to determine how central tendency is extracted? We focused on the potential role of visible persistence mechanisms that support temporal integration. We tested whether mean orientation could be accurately recalled from brief visual displays using the successive field task. On separate blocks of trials, participants were asked to report the location of a split element (requiring differentiation of frames), a missing element (requiring integration across frames), and the average orientation of elements pooled across both frames (central tendency recall). Results replicate the expected tradeoff between differentiation and integration performance across inter-frame interval (IFI). In contrast, precision of mean estimates was high and invariant across IFIs. A manipulation of within-frame distributional similarity coupled with simulations using 12 models supported 2-item subsampling. The results argue against the “strategic” interpretation of subsampling since 2-item readout was predicted by information theoretic estimates of STM encoding rate: the 2 items were not from a superset in STM. Most crucially, the results argue against the various early “preattentive/parallel/global pooling” accounts and instead suggest that non-selective readout of information from iconic memory supplies a relatively small amount of item information to STM, and it is only at this point that the computation of ensemble averages begins.

**Keywords** Attention: Interactions with Memory · Visual working memory · Short term memory

## Introduction

In the introduction to his classic 1960 monograph, George Sperling writes that when arrays of stimuli are presented tachistoscopically “observers enigmatically insist that they have seen more than they can remember afterwards, that is, report afterwards,” (p. 1). Sperling took this as an initial clue that there may be a brief form of visual storage preceding short-term memory in the processing stream. What is particularly interesting about this statement today, however, is what it reveals about the representational

layers or components that survive iconic decay. Specifically, participants appear to have statistical information about the items that occupied the icon even when those items can no longer be remembered, the statistical information in this case being the sample size.

The following year, another now-classic paper on iconic memory was published, this time by Averbach and Coriell. Reaching conclusions similar to those of Sperling (1960) regarding the capacity and timecourse of iconic memory, the authors also demonstrated the important role of masking stimuli in controlling readout from the sensory store. They documented an effect whereby a circular mask “erases” the visual percept of a preceding stimulus falling at its center when target-mask SOAs were below the estimated lifetime of the icon. These results inspired many subsequent studies which used masking manipulations to determine in detail the kind of information retained from brief displays. Equally interesting, however, is the conclusion Averbach and Coriell (1961) reached with respect to near-simultaneous target-mask onset conditions with the circular mask: “High

---

✉ Chad Dubé  
chaddube@usf.edu

<sup>1</sup> University of South Florida, Tampa, FL, USA

<sup>2</sup> University of Trento, Trento, Italy

<sup>3</sup> New York University Abu Dhabi, Abu Dhabi,  
United Arab Emirates

performance when the circle follows immediately after the array is due to simple temporal averaging in the visual system. This results in array and circle being effectively superimposed, which does not significantly affect legibility,” (p. 321). A second experiment comparing iconic memory performance following the circular mask to performance following a grid mask showed a large decrease in partial report performance for the grid mask, consistent with the authors’ interpretation.

The idea put forth by Averbach and Coriell was essentially that items falling within a temporal window contained within the lifetime of iconic memory are summed, superimposed, or averaged together, while those that straddle contiguous “windows” will interact, leading to erasure of percepts formed or initiated in the earlier window by those of the later window. In other words, mechanisms of *integration* and *differentiation* operate on iconic memory contents. Items falling at greater temporal distances from one another produce typical perception of succession. Once again, statistical information appears to be necessary, and this time the statistic in question is a sum or average.

In what follows, we evaluate the notion that at least some forms of central tendency representation in visual short-term memory may be extracted pre-attentively, rapidly, and/or automatically. We conclude that this suggests central tendency information may be extracted directly from iconic memory. Second, we identify a particular iconic memory subprocess, temporal integration, that appears to be a plausible candidate mechanism for carrying out or contributing to the early extraction. Finally, we report an experimental test of both of these notions.

## Rapid, pre-attentive averaging?

There is now a vast and growing literature on ensemble coding in visual perception and short-term memory (VSTM). This literature suggests that participants can extract statistical summaries such as the central tendency of feature values over ensembles of similar stimuli. In particular, the literature contains evidence that central tendency representation may occur rapidly, pre-attentively, and/or automatically (Alvarez & Oliva, 2007, 2008, 2009; Chong & Treisman, 2003, 2005a, b, 2008; Emmanouil & Treisman, 2008; Oriet & Brand, 2013); but see De Fockert and Marchant (2008), Myczek and Simons (2008), and Whiting and Oriet (2011).

For instance, Alvarez and Oliva (2008) used a motion tracking task to investigate participants’ abilities to report their memory for the mean location of distractors. The display consisted of eight moving circles, four of which were designated as targets. The participant’s task was to track and report the number of instances in which the

targets crossed boundary lines in the display. Following this, a subset of the distractors was presented in rest at their final positions from the trial. The participant was asked to report the center of mass corresponding to the remembered position of the missing distractors. Participants were as accurate in this judgment as they were in control trials in which the same centroid judgment task was applied to targets. The authors interpreted these results as showing statistical extraction for items that were unattended. Further work by Alvarez and colleagues, using similar tasks, has provided support for this initial conclusion (see, e.g., Alvarez & Oliva, 2009).

“Unattended”, of course, is not necessarily equivalent to “pre-attentive.” Chong and Treisman (2003) provided evidence of an early, parallel, automatic, and pre-attentive extraction of central tendency. In their view, the pre-attentive state involves “distributed” or diffuse attention Chong & Treisman (2005a, b), a state existing prior to attentional selection or “focused” attention.<sup>1</sup> The extraction of central tendency, which the authors described as “automatic”, may be aided by this state.

Participants were shown left/right displays of circles varying in size according to a fixed distributional form. Experimental conditions included ensembles of heterogeneous size, homogeneous size, and single items (one item on each side of the display). The authors varied stimulus duration (from 50 to 1000ms, with no backward mask) as well as whether the left and right displays were simultaneous or successively shown, and the ISI between displays in the latter condition. Participants were asked to report which side contained a larger mean size (or circle size if there was one item). Performance was measured in terms of the threshold level of difference in mean size across the left and right displays necessary to maintain 75% correct performance.

The key finding from the experiments by Chong and Treisman is that mean discrimination performance for simultaneous conditions was highly accurate at the earliest, 50ms stimulus duration and accuracy did not change with duration. An experiment varying the distributional form from which sizes were sampled as well as the combination of distributions across the left and right displays showed only a 2% increase in thresholds for the most dissimilar distributions. The authors concluded that the results rule out mental arithmetic and subsampling as alternative explanations, and concluded that mean representation

<sup>1</sup>Note that this definition of “pre-attentive,” while consistent with that of Treisman and Gormican (1988), appears to differ from other definitions of the term in the literature. For instance, Treisman (1985) writes “Some discriminations appear to be made automatically, without attention and spatially in parallel across the visual field. Other visual operations require focused attention and can be performed only serially.” See also Treisman et al. (1992).

occurs automatically and in parallel across the display (see also Chong and Treisman, 2005a, b, 2008).

In a more recent study, Whiting and Oriet (2011) pointed out that the arrays used in Chong and Treisman's work were unmasked. Using a size averaging task, Whiting and Oriet found that when the arrays were masked at short SOAs, performance on the task declined and began to show dependence on estimates from prior trials. Unlike the earlier study (Chong & Treisman, 2003), the presence of a backward mask limited the duration of iconic memory. Without a mask, information from iconic memory would continue to persist and be stored in VSTM. Since backward pattern masking is known to disrupt iconic memory transfer to STM (the amount of information that can be non-selectively read out) but has little to no effect on information already in VSTM storage (Coltheart, 1980; Gegenfurtner & Sperling, 1993), these results argue against the pre-attentive view and in favor of one in which averages are computed on the basis of VSTM contents. Put another way, if in fact means were extracted directly from the iconic store, a subsequent mask would have no effect on mean estimates, all else (e.g. lags) being equal, since the full iconic store would have contributed to the estimate, not just the few items surviving the readout process (which is disrupted by the mask). Our work, reported later, further examines this hypothesis about the processing locus of averaging.

The results from Chong and Treisman suggest an early mechanism operating on information in iconic memory, one which could integrate information within each display separately to allow a subsequent "differencing" or discrimination mechanism to compare them. In contrast, the effect of the SOA when stimuli were masked (Whiting & Oriet, 2011) is indicative of ensemble computations on the contents of VSTM, rather than iconic memory. This suggests a consideration of the rather extensive early literature on integration and segregation operations on the contents of sensory memory. We turn to this next.

## Integration and segregation mechanisms in iconic memory

The early notions regarding the aggregation, interaction, and separation or differentiation of stimuli in time are the result of decades of earlier research stretching back at least to Wundt (1900). As discussed in Sperling (1960), Wundt's question was whether the true duration of a sensation or percept (controlling for retinal after-images) was coterminous with the physical stimulus. Wundt cited early "two-flash" flicker fusion research conducted by Weyer (1899) as evidence in favor of persistence of vision beyond the physical stimulus duration and distinct from retinal afterimages.

Modern cognitive psychology, starting at least as early as Sperling (1960), detailed the existence of a second, informational component. Informational persistence may be thought of as a non-visual, symbolic or "coded" representation of the stimuli in an array. Decades of work establishes this fundamental visible/informational difference in underlying variables, showing that, unlike in partial report tasks, tasks (such as temporal integration tasks) that are believed to rely on visible persistence are disrupted when the eyes move between frames (Irwin, 1991), that memory for stimulus location decays more quickly than memory for the identity of the stimulus (Dick, 1969; Mewhort et al., 1981), and, most strikingly, that tasks thought to measure visible persistence show an "inverse duration effect" (persistence decreases with stimulus duration) while the partial report task does not (performance increases with duration; Coltheart, 1980; Sperling, 1967). Hence, visible persistence, along with the "informational persistence" which partial-report techniques are intended to measure, is considered today to be one of two key components of iconic memory (see Coltheart 1980 for an early, but comprehensive, review).

Visible persistence research in the 1950s and 1960s directed a great deal of effort toward the measurement of a hypothesized basic unit of duration for simple percepts. This hypothetical unit was referred to as the "perceptual moment" (Stroud, 1956). A variety of techniques were adopted, including persistence of form (Haber & Standing, 1969), temporal numerosity judgments (White, 1963), judgments of synchrony and duration (Efron, 1970), and successive-field techniques (Eriksen & Collins, 1967; Hogben & Lollo, 1974). Though many of these studies suffer from confounds related to possible response criterion artifacts (Long, 1980), they nonetheless show a respectable level of consistency in their estimation of visible persistence's duration. This value, usually in the range of 80-200ms, is reasonably close to the estimates of "iconic memory" lifespan derived from the partial-report and delay-of-masking techniques inspired by Sperling's work.

Of these tasks, the successive-field procedure is of particular interest to our current work. The procedure, pioneered by Eriksen and Collins (1967), involves the presentation of two stimulus frames in rapid succession and separated by a blank frame. When superimposed, a coherent image is produced. By varying the duration of each frame and the inter-frame interval (IFI, duration of the blank screen) and measuring performance in detecting the composite image, one can in principle measure the duration of perceptual integration windows or "moments."

In an interesting twist on the original procedure, Hogben and Lollo (1974) presented participants with frames containing dots arranged in matrices. Superimposed, the frames share one empty cell. The participant's task was

to report the position of the missing element (circle). Successful performance in the task therefore requires integration of the information presented in the two displays. In line with prior work, Hogben and DiLollo found that participants' accuracy in the task was high for short IFIs and fell off rapidly with increasing IFI. An important result was that participants reported becoming aware of distinct percepts around an IFI of 30–50ms. Manipulations of the number of frames and distribution of IFI levels within trials led the authors to conclude that two mechanisms operate within the timescale of iconic memory to i) integrate information within a brief window of approximately 120ms and ii) to segregate information falling within adjacent windows of time.

A recent study by Wutz et al. (2016) used a variant of the successive field procedure that contained both a missing element and a split element in which one element in the display was split across the two frames (shown one half at a time). On integration trials, participants reported the location of the missing element, as in Hogben and Lollo (1974). On segregation trials, participants reported the location of the split element. In line with much prior research, they found a clear crossover interaction between task and IFI: integration performance decreased linearly with IFI and segregation performance increased with IFI.

Next, we report an experiment using a variation of the successive-field task, building on the task used by Wutz and colleagues. Perhaps the most important addition we make is the addition, on some trials, of a requirement for participants to report the average orientation of the stimuli pooled across the two frames. As the trials are identical across the integration and averaging tasks with the exception of the judgment required, the design allows us to determine whether central tendency can be extracted from the iconic store or instead involves operations on VSTM. We also investigate whether temporal integration mechanisms may be involved.

In detail, our design uses an array of 16 items distributed across two very briefly presented frames. Since this number is beyond the putative capacity of VSTM (Cowan, 2001) to store, accurate performance on the averaging task would support an iconic memory contribution. This, of course, assumes that participants do not engage in a strategy of “subsampling” the displays (Myczek & Simons, 2008; De Fockert & Marchant, 2008). So, we also include a parametric manipulation of within-frame variance in order to test for subsampling. Due to the rapid temporal characteristics of the design, intentional/strategic subsampling would be improbable as saccades during presentation of the frames produce the perception of a “blur” of items and there is very limited time to move visual fixation. If, on the other hand, central tendency representations are generated from VSTM, a necessary

“subsampling” of items would be expected due to normal limits on transfer of information to VSTM. Finally, if iconic contributions operate through temporal integration mechanisms, we should see parallel decreases in accuracy of both integration and averaging performance with increases in the interval between frames.

## Methods

### Participants

Ten University of South Florida undergraduate students were recruited from the university subject pool, ranging in age from 18–22 years old (eight female, mean age = 19 years, SD = 1.34 years.) All had normal or corrected-to-normal vision. A sample of 20 was originally planned, but we closed the lab midway through data collection on March 2, 2020 in response to reports of the COVID-19 outbreak. Subjects completed the experiment in one session and received course credit as compensation for their time. All experimental procedures were approved by the University of South Florida IRB.

### Apparatus and stimuli

Stimuli were created using the Matlab Psychophysics Toolbox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007), and were presented on a 1920 x 1080 pixel resolution 25” LCD monitor with a 240 Hz refresh rate. The monitor was tied to a PC that was custom built to optimize timing and processing speed, and to minimize lags. The PC's frame rate was 125 fps, and was benchmarked in FRAPS (frame rate benchmarking software; (Beepa, 2013)) during test runs of the experiment to verify observed frame durations matched those specified in Matlab. Participants were seated approximately 100 cm from the screen, making each pixel of the screen approximately 0.02° of visual angle. Responses were recorded via mouse movement and button clicks that corresponded to the current task.

The stimuli consisted of black annuli (each subtending 0.5° of visual angle in diameter) presented within an invisible 4x4 grid of equally-spaced possible presentation locations. The invisible grid was positioned centrally on the screen. There were 3.5° of visual angle in horizontal and vertical size and 0.5° of space between grid element locations, measured center to center.

The 16 possible stimulus locations within the grid were randomly divided into two frame presentations, split by a blank frame, each consisting of 8 positions designated to each frame. Each annulus contained a 45° gap in its ring, and the gaps were presented at either a 0°, 45°, 90°, or 135° orientation. All stimuli were presented against a uniform

gray background, set at 50 percent of the monitor's RGB range (i.e. at Weber contrast). While the current stimuli are not common in psychophysical orientation research, work in visual acuity has found that contrast acuity for Landolt C is about double that of sinusoidal gratings and well above the limit of acuity for the eye (see Bondarko & Danilova, 1997; McAnany & Alexander, 2008), and that orientation discrimination is relatively high (Harrison & Bex, 2015), which suggests that these stimuli and arrangement should be more than optimal for use in the current study.

For the target stimulus of the segregation blocks, one randomly selected location of the 16 presentation locations was occupied by an annulus that was bisected into two halves, dividing the gap in the ring centrally; one piece of this annulus was presented in the first frame, and one half was presented in the second frame, such that overlaying the two halves would create a full annulus object. Identification of this stimulus from the distractors requires the participant to represent each frame distinctly within time, and to identify the odd element.

For the target stimulus of the integration blocks, one randomly selected position out of the remaining 15 stimulus presentation grid locations consisted of no annulus on either frame, and was therefore missing an element in the position. Identification of this missing element from the distractor stimuli requires the participant to represent an integrated percept of both frames to identify the location where no information was presented on either frame.

Of the 16 possible stimulus locations in the integration and segregation task blocks, 14 locations consisted of full annuli (each with a 45° gap in the ring), 1 location consisted of an annulus segmented across the two frames, and one location where no annulus was presented, equating to 7.5 annuli being presented per frame in these tasks.

In the orientation averaging task, all 16 locations across both frames (8 locations per frame) were occupied by an annulus object with a 45° gap cut out of the ring. Annuli in the orientation averaging and integration trials contained only one 45° gap missing from the ring, but within the segregation blocks a second 45° gap 180° from the first was added. This extra gap was added for these trials due to motion cues of the edges of the ring making the split element too easy to identify, relative to the distractors. As compared to more conventional Gabor patch stimuli, this stimulus design eliminates the ambiguity of the direction of the orientation versus its 180° reflection by having a single orientation angle denoting gap. Therefore, subjects should not need to resolve whether the stimuli was oriented at, for example, 90° or 270°.

To investigate the possibility of sub-sampling strategies, distributional properties of each frame were independently manipulated. Table 1 lists the full set of stimulus parameters to be detailed in this section. Within each block, each

frame was set to one of 4 predetermined levels of variation in stimulus orientation within the frame ( $\sigma_c^2 = 0, 0.25, 0.39, 0.46$ , where  $\sigma_c^2$  denotes the circular variance). The number of presentations for each of these variance levels was equivalent within each task and within each frame, and the order of assignments to a frame was random across trials. As this random ordering was carried out separately for each frame, the particular combinations of variances for frames 1 and 2 were also random (though, again, an equal number of variance presentations for each variance was preserved for each frame). Further breakdowns of the distribution of stimuli parameter permutations is detailed within the *Stimulus Distributions* section<sup>2</sup> of the [Appendix](#).

## Procedure

The procedure is illustrated in Fig. 1. Each stimulus display was presented centrally. Trials began with a 500ms fixation cross subtending 1° of visual angle in size in the horizontal and vertical directions. This was followed by a blank screen whose duration was sampled randomly on each trial from a uniform distribution with a range between 500ms and 1500ms, in steps of 10ms. Afterwards, the two frames were presented for 8ms each, divided temporally by a blank inter-frame interval that ranged in duration between 30ms–86ms in steps of 8ms. These inter-frame intervals were presented on equal numbers of trials within each block, but were presented in a random order. Subjects received a response screen 500ms after the offset of the second frame, within which they had unlimited time to respond. Before beginning the experiment, participants completed sets of 13 practice trials to gain familiarity with each task and to ensure comprehension of instructions. The first 3 practice trials were presented with extended frame presentation durations (200ms each frame) to give the participants a chance to become familiar with the stimuli and the mode of their presentation. The following 10 practice trials were presented with the same frame duration structure as the actual experimental trials. Participants received feedback and were monitored by a researcher while completing these initial tasks to ensure appropriate performance. If a subject's performance on the practice trials was too poor then an additional round of practice trials would have been performed for said subject, though this was not necessary for any of the tested subjects.

<sup>2</sup>It could be argued that the presence of repeated angles in the arrays leads, functionally, to lower set sizes than the 16 we assume are subjected to encoding. However, the recent literature on this point shows that averaging efficiency is comparable for sets containing repetitions and those that do not, when range confounds are controlled in the comparison (Marchant et al., 2013; Tiurina & Utochkin, 2014).

**Table 1** Possible stimulus angle configurations for a frame

Angle Sets (in degrees)								$\mu_c$	$\sigma_c^2$
0	0	0	0	0	0	0	0	0	0
45	45	45	45	45	45	45	45	45	0
90	90	90	90	90	90	90	90	90	0
135	135	135	135	135	135	135	135	135	0
0	0	0	0	45	90	90	90	38	0.25
0	0	0	45	90	90	90	90	52	0.25
45	45	45	45	90	135	135	135	83	0.25
45	45	45	90	135	135	135	135	97	0.25
0	0	0	0	45	45	135	135	35	0.39
0	0	0	45	45	90	135	135	52	0.39
0	0	45	90	90	135	135	135	83	0.39
0	0	90	90	135	135	135	135	100	0.39
0	0	0	0	0	90	135	135	34	0.46
0	0	0	0	90	90	135	135	53	0.46
0	0	45	45	135	135	135	135	82	0.46
0	0	45	135	135	135	135	135	101	0.46

Each row of the table was assigned the same number of times (though in a random order across trials) to the frame in the experimental trials, and orientations within a given row were randomly assigned to locations within the frame. This logic was followed for each frame, meaning the specific combination of means (and variances) for frames 1 and 2 were randomly determined on each trial. See Text for further details

Participants completed 4 repetitions of blocks consisting of each individual task, for a total of 12 blocks in the experiment. The order of the tasks within each block was semi-random, with a restriction placed that two of the same task could not occur in succession. Within each individual task block, instruction of the current task was given before the first trial, followed by 60 trials of the instructed task. After one block of each task was completed, participants were instructed to take a 2-minute break before beginning the next set of task blocks. Thus, in total there were 240 experimental trials devoted to each task, and a total of 720 critical trials in the experiment.

## Response tasks

There were three tasks, each carried out in separate blocks of trials, within-subjects: *Integration*, *Segregation*, and *Averaging*. For each task, participants made the relevant response via mouse position movement and mouse button clicking to submit a response. A schematic of the tasks is shown in Fig. 1. The response screen consisted of a 4x4 grid of the numbers 1-16 presented in the same locations of the 4x4 grid of possible stimulus positions. The mouse cursor was analogous in design to a stimulus annulus (black, 0.5° visual angle in size, and a 45° gap in the perimeter), and would start at a randomly selected orientation between 0° and 180°, independent of the previously presented stimuli or correct response. Responses were restricted to the range of 0° and 180° for the orientation task, as all possible correct

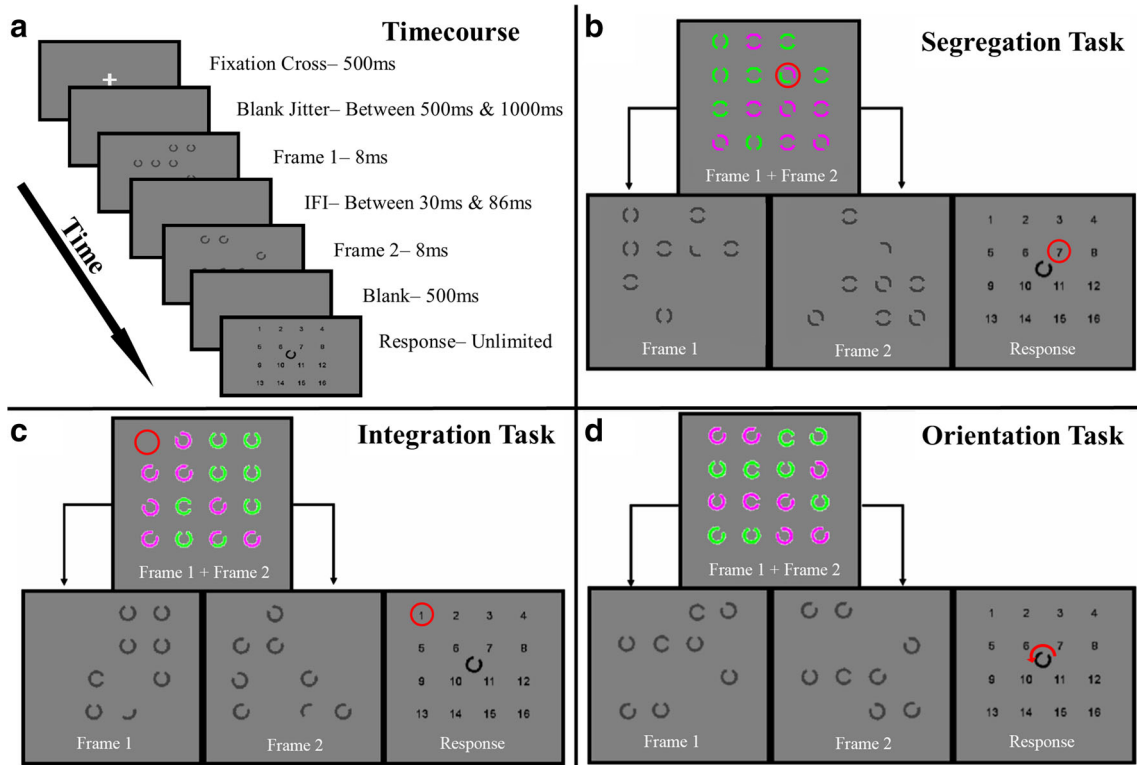
mean values occurred between the endpoints of 0° and 180°.

In the integration and segregation tasks, the cursor was initialized at a random location within the imaginary 4x4 grid. Participants then moved and clicked the cursor via the mouse to the number situated in the grid position that corresponded to the missing or segmented element, respectively. In the orientation averaging task, the annulus was fixed to the center of the imaginary grid. Moving the mouse forward and backwards rotated the annulus, and participants were instructed to click the mouse when the probe annulus was oriented to the average orientation of the 16 annuli presented in the preceding 2 stimulus frames.

## Results

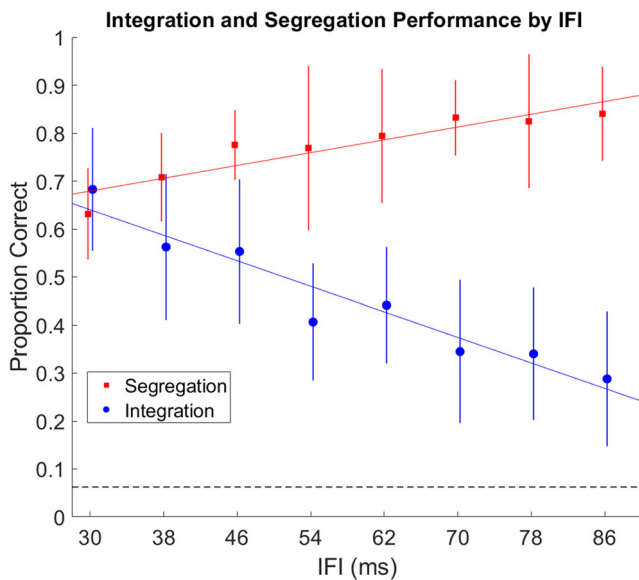
### Integration and segregation tasks

The results of the temporal integration and segregation tasks are displayed in Fig. 2. A simple linear regression analysis conducted on the group data in the Figure revealed that the IFI value accounts for 84.1% of performance variation within the segregation task ( $\beta = 0.0033$ ), and 92.6% of performance variation within the integration task ( $\beta = -0.0066$ ). The same analysis was also applied to each individual participant's data. Those results are shown in Table 2. As can be seen in the Table, although the quality of the fits is variable, the individual participants are for the most part consistent with the fits to the group data.



**Fig. 1** Experimental timeline (A) and example stimulus displays for each task (B,C,D). Each task followed the same timecourse as shown in (A), with only the goal and response of the subject varying across tasks. The green and purple colored stimuli in the “Frame 1 + Frame 2” overlay sections of B, C, and D represent the items from frame 1

and frame 2, respectively. The red circles in (B) and (C) represent the target of the task, to which the participant should move the mouse to click. The red arrow in (C) represents the rotation of the central probe for the averaging response



**Fig. 2** Performance on the integration and segregation tasks by the inter-frame interval duration, aggregated across all subjects. The bars extending from each point represent the point’s 95% confidence interval, and the probability of guessing is plotted as a black dashed line. Integration performance is negatively correlated with IFI, whereas segregation performance is positively correlated with IFI

All participants showed positive slopes in the segregation condition, and negative slopes in the integration condition, and the means of those slopes (.0034 and -.0066 for segregation and integration, respectively) are very similar to the parameters from the group fits already mentioned.

**Table 2** Regression results from fits to individual-participant data from the Integration and Segregation tasks, and column means

ID	Integration			Segregation		
	$\beta_0$	$\beta_1$	$R^2$	$\beta_0$	$\beta_1$	$R^2$
1	.7138	-.0069	.4672	.7747	.0021	.3311
2	.5638	-.0050	.6409	.6891	.0035	.3934
3	.8558	-.0093	.7813	.6312	.0024	.3207
4	.8199	-.0064	.7149	.5695	.0045	.8490
5	.7969	-.0062	.3756	.4344	.0023	.4999
6	1.2187	-.0064	.8023	.4267	.0008	.0098
7	.6678	-.0038	.3348	.6449	.0033	.6984
8	.9814	-.0088	.7168	.6913	.0032	.5082
9	.9121	-.0057	.5320	.4645	.0052	.7517
10	.8460	-.0079	.8171	.4562	.0062	.6172
Mean	.8376	-.0066	.6182	.5782	.0034	.4979

A two-way repeated measures ANOVA was conducted to determine the effect of task type and IFI level on performance (percent correct). The analysis found a significant interaction between the task and the IFI level,  $F(7, 63) = 24.614$ ,  $p < 0.001$ ,  $\eta_p^2 = 0.732$ , which can be clearly seen in Fig. 2. Simple main effects analysis showed that performance on the segregation task was significantly better than the integration task for the 46ms, 54ms, 62ms, 70ms, 78ms, and 86ms IFI levels ( $p < 0.05$  for each of the levels), while performance on the 30ms and 38ms levels did not significantly differ between the tasks ( $p = 0.519$ ,  $p = 0.177$  respectively).

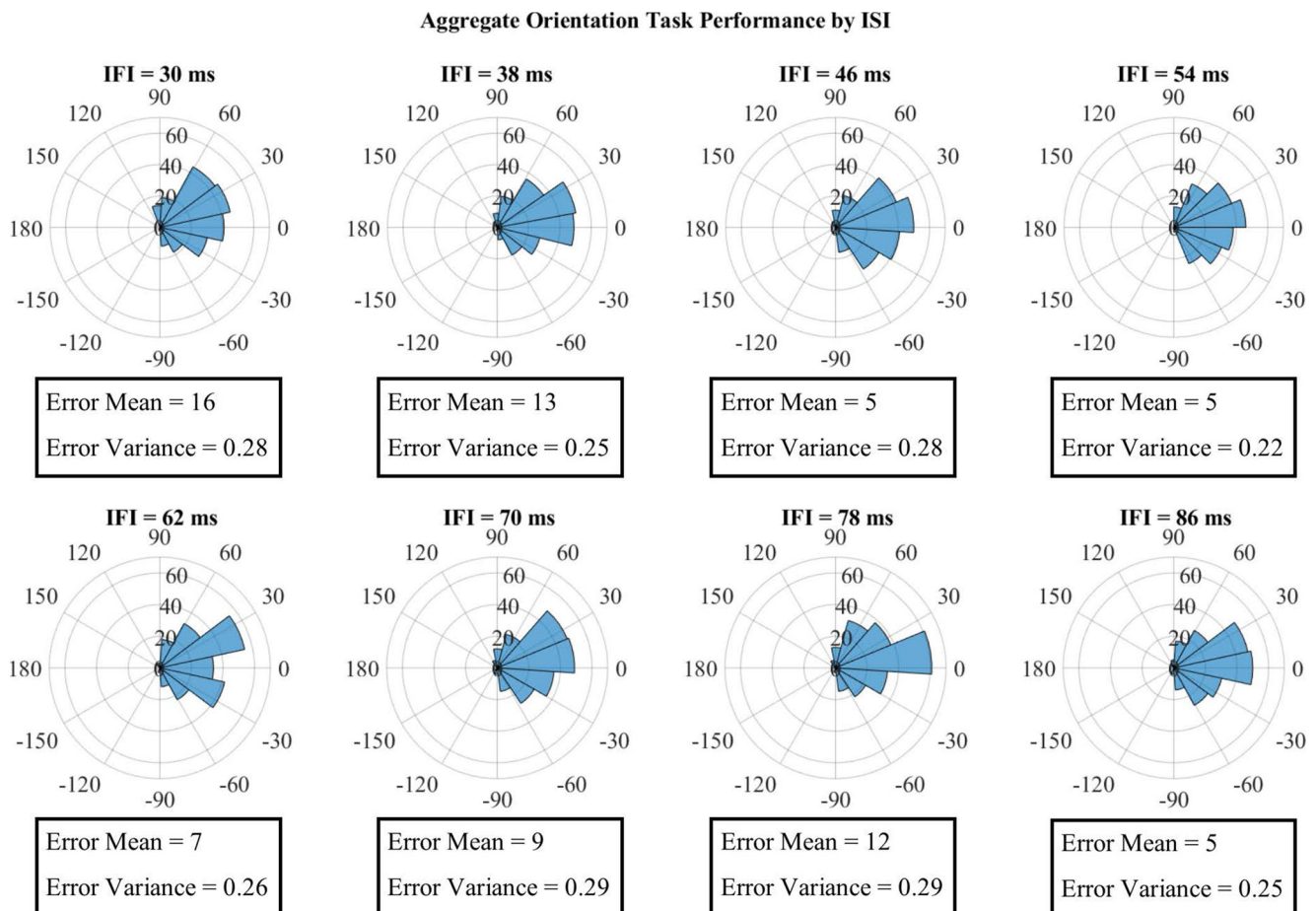
Comparing these results to those of Wutz et al. (2016 [Figure 1B]), we see that despite very similar design and task parameters, the crossover point between integration and segregation performance occurs at a much lower IFI in the current study (30ms vs. ~60ms in Wutz et al.). In pilot work, we found that segregation performance was affected by the visual angle subtended by the stimulus display. However, this did not impact integration or averaging performance, which showed results similar to those reported here. Most

importantly, however, the basic finding of a tradeoff in task performance as a function of IFI is very similar to what was reported by Wutz et al. and sensible in light of the prior work using successive field tasks reviewed earlier.

Having established reliable and robust integration effects across IFI levels, we now turn to our main question: Does orientation averaging across the two frames show a systematic decrease in accuracy as IFI is increased? Such a pattern would suggest that visible persistence and temporal integration mechanisms are involved in ensemble coding within the iconic store. If, however, accurate mean estimates persist unchanged with IFI, this would leave informational persistence as the sole mechanism responsible for providing the raw materials for mean extraction via iconic memory.

### Orientation averaging task

The distributions of response errors (participant estimate - true mean) for each level of IFI, aggregated across participants, are displayed in Fig. 3. The precision (1/variance) of the response error distributions will be



**Fig. 3** Circular histograms of aggregated participant response errors (estimate - true mean) on the orientation averaging task, by the inter-frame interval duration. Mean error values closer to zero represent responses that did not differ much from the true circular average of the set of stimulus angles



the focus of analysis for the orientation task. Increases in precision are driven by subjects making judgements of the average orientation that have a similar degree of error from the true mean of the stimuli set. If subjects were unable to complete the task correctly or were guessing often for responses, the variance would be high and precision would be low. Although it is clear that there is an overall positive bias in the error distribution means, we are uncertain of the origin of this bias. We speculate it may reflect a motor perseveration error, as counter-clockwise probe movements corresponded to forward mouse movement. As the present analysis's interest is the subjects' ability to accurately recreate the central tendency representation, and not the additive effect of the positive bias displayed by the error distribution means, the precision of the error distributions give the greatest amount of pertinent information. Regardless, a Watson-Williams test for the equivalence of circular means shows no significant difference of mean error across IFI levels,  $F(7, 1792) = 1.88$ ,  $p = 0.070$ .

Due to the circular nature of the dependent variable measure, a Rayleigh Z-test for uniformity was performed on the error distributions at each level of IFI (Fisher, 1993). Results for each level are shown in Table 3. They indicate that the distributions are indeed not uniform. This indicates that participants were rarely, if ever, guessing the response. A Bartlett's test for the equality of variances was conducted to examine the effect of IFI value on orientation averaging error variance (Zar, 1999). No significant difference was found in participants' error variance for the levels of IFIs,  $\chi^2(7) = 6.992$ ,  $p = 0.430$ . These results appear to rule out temporal integration as a key ensemble coding mechanism within iconic memory.

## Subsampling

Studies of ensemble coding raise the question of whether participants might use control processes to average a small sample of items from the full displays. For instance, Myczek and Simons (2008) demonstrated via simulations that several of the results reported in the literature, such as those of Chong & Treisman (2003, 2005a, b) and Ariely (2001), could be reproduced by sampling only one or two items. A related paper by De Fockert and Marchant (2008) demonstrated that if an item in a briefly-displayed ensemble

of circles is cued (via its size or color) for a later judgment regarding the cued feature (e.g. whether the circle was small or large), subsequent estimates of the average circle size are biased in the direction of the cued circle's size. The authors point out that the results are equally consistent with a subsampling strategy as well as with a weighted average, in which weights might be assigned based on the fidelity of the cued item's memory representation (see Tong et al., 2019, for a similar proposal). Finally, a study by Attarha et al. (2016), using dynamic dot motion and size techniques pioneered by Albrecht and Scholl (2010), suggested that a temporal integration window may produce subsamples that support averaging.

The initial work by Myczek and Simons prompted a series of replies. Among these was one from Chong et al. (2008) that provided new data arguing against the subsampling position (for recent evidence from CDA amplitudes supporting Chong and Treisman's position, see Baijal et al., 2013). In a response to the various critiques, Simons and Myczek (2008) pointed out that their claims only applied to object size. In their view, the idea that participants extract averages of other features such as orientation (the one used in the present work) is "uncontroversial" (p. 1335). Additionally, as pointed out by Whitney and Leib (2018), a subsampling strategy does not rule out an averaging mechanism so long as more than one item is sampled. Finally, the current study uses an array of 16 items, which is beyond the putative VSTM capacity limit of 4 items (Cowan, 2001). This means that if in fact control processes were operating, the contents of VSTM to which they were directed would have to have been a subsample. Hence, conclusions about subsampling may entail conclusions about which memory store is involved in our task (the first of our two research questions). If participants are not subsampling, they cannot be deriving their estimates from VSTM storage.

For these reasons, we included an analysis and manipulation to determine whether samples of one item from one frame could reproduce our results. To accomplish this, the variance of the set of stimulus orientations within each frame was independently fixed to one of 4 levels across trials (see *Apparatus and Stimuli*). A strategy of sampling only 1 item leads to a clear set of predictions in this design. First, if participants always sample from frame 1 or frame 2, then we should see a main effect of either frame 1 or frame 2. Specifically, for a given frame, errors should increase (precision decrease) with higher intra-frame variance because the sampled item will in general be further from the overall mean across frames (see Table 1). Assuming, on the other hand, that participants switch frames across trials or across subjects, main effects of both frames should be evident. Finally, if participants sample a small proportion of the items from only one frame, a distinct trade-off can be predicted.

**Table 3** Results of Rayleigh Z-tests of uniformity, at each IFI level

IFI (ms)	30	38	46	54	62	70	78	86
Z	118	127	116	139	124	111	111	125

All  $p < .01$

To understand this, consider a case where participants only sample 2-3 items from frame 2. Error will of course increase with the intra-frame variance (IFV) of frame 2. However, the random combination of IFV between frames means that, for the highest level of IFV on frame 2, the IFV of frame 1 cannot exceed IFV of frame 2 and on average will be lower than IFV of frame 2. This means that, if participants subsample from frame 2, the error conditional on frame 1 should decrease with IFV as the error conditional on frame 2 increases with IFV. The opposite pattern (increase with IFV on frame 1 and decrease on frame 2) should occur if participants subsample from frame 1. These predictions thus offer a clear test of recency-based sampling and integration processes recently proposed by Tong et al. (2019) and others. We also verify that these predictions do in fact follow logically from our design, using subsampling models.

We first turn to an analysis of the participants' error variances, shown in Table 4. Several Rao's tests for equality of variation were conducted to determine the effect of intraframe variance level on participants' variation in response error. For frame 1, these results show a significant effect of IFV on participants' error variance,  $H(3) = 8.580$ ,  $p < 0.05$ . Pairwise comparisons reveal that this difference is driven by a peak in error variance at the lowest IFV level. For frame 2, we once again find a significant effect of IFV on participants' error variance,  $H(3) = 78.718$ ,  $p < 0.05$ . The variation in response errors increases with IFV level, until a plateau at the largest IFV levels.

Together, these results show that participants' error variance is systematically increasing with frame 2 IFV, but not with frame 1 IFV, which actually shows a hint of the opposite pattern. This crossover pattern in error variation is what one would expect if subjects were subsampling primarily from frame 2.

Having established that participants do in fact seem to be subsampling from frame 2, we turn to simulations in order to i) verify that our predicted (and observed) trade-off

does indeed follow from frame 2 subsampling and ii) determine how many items are being sampled. A separate set of simulated responses was constructed by replacing the subject's response in the experimental data set with the average of a subset of randomly sampled angles from the 16 angles used within the experiment trials. The subset average became the new response, and the difference of the simulated response and the true mean of the 16 item angles became the new response error. This method allows for the preservation of the distributional properties of the original data set.

The results are shown in Fig. 4. Since the data are being described by models with zero free parameters, with all models assuming perfect estimation responses for the sampled item(s), all variances of the observed data are generally underestimated in the models by a constant. Since the patterns of changes with IFV are what are crucial to tests of these models, we mean centered the data, and mean centered the model, separately for each frame before plotting them.

As is clear in Fig. 4, the 2-item models are superior to the other models. It should be noted that these models assumed ideal, noiseless encoding, which provides a particularly stringent criterion for the models to pass. The good fit to the data by the 2-item models is therefore doubly impressive. Comparing these best-fitting models, we find that the Random 2-Item model and frame 2 2-Item model were essentially tied. The closest competitor, the 4-Item Random model, differed greatly in fit. The two worst models were the 1-item models, both of which produced large negative  $R^2$  values due to the residuals being greater than the variability of the observed data, calling into question the wisdom of referring to the statistic  $1 - \frac{SSR}{SST}$  as a squared value.

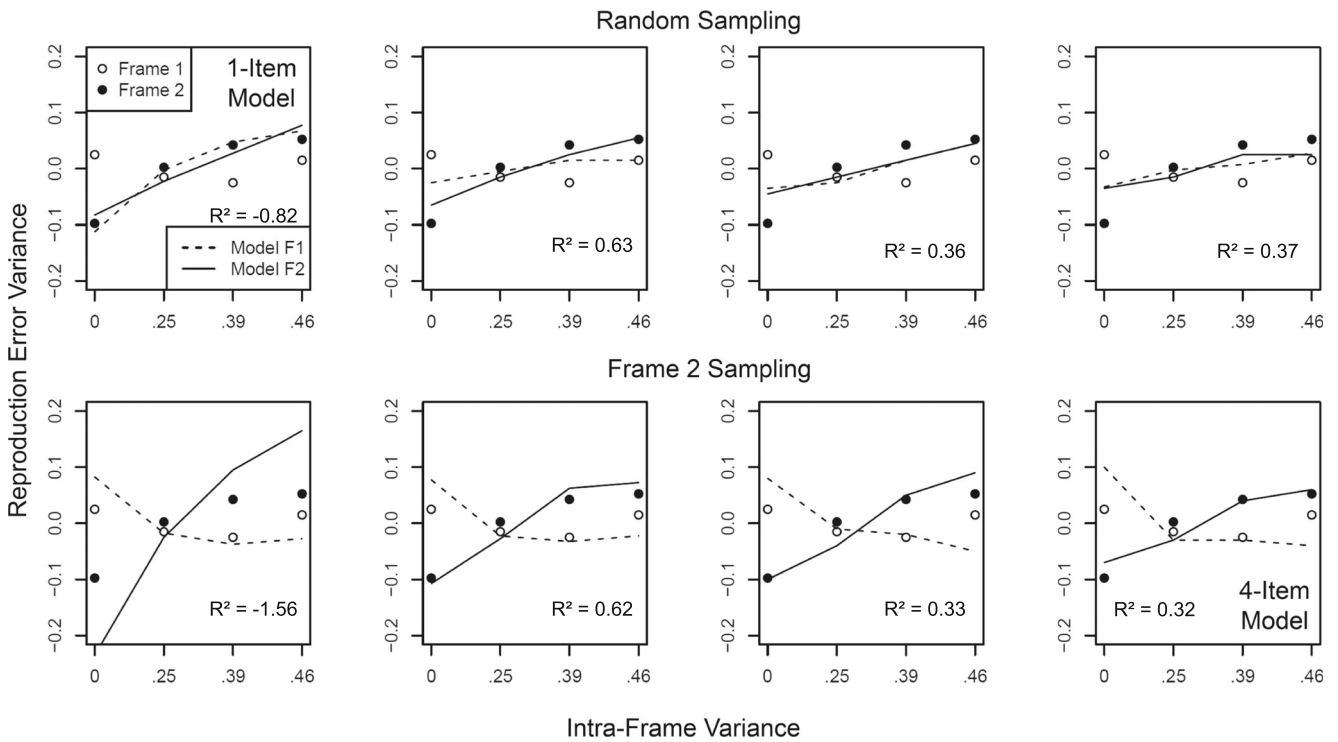
Finally, we note that all frame 1 models departed greatly from the observed data. These models produced the opposite of the crossover pattern exhibited by the frame 2 models, as predicted above, and thus have been omitted from further discussion.

It is also possible that the subsampling behavior may be dependent on IFI, such that temporal integration may modulate the recency effect of frame 2 at the lower IFI levels, and therefore facilitate random subsampling from both frames. To analyze this possible strategic modulation, the effect of each frame's IFV was further broken down by IFI. As can be seen in Table 5, subject error variance significantly increases with increases in IFV of frame 2. This effect holds at nearly every level of IFI, although the effect is smaller and non-significant for the two shortest IFIs. This is consistent with the idea that random 2-item sampling may occur at the point of maximal integration of frames, while a recency effect (frame 2 subsampling) emerges as the two frames are perceptually segregated (Fig. 5).

**Table 4** Group-level circular error variances (1/precision) of participants' estimates of mean orientation as a function of the intra-frame orientation variance for frames 1 and 2

	$\sigma_c^2$			
	0	.25	.39	.46
Frame 1	.29	.25	.24	.28
Frame 2	.17	.27	.31	.32

The results show a tradeoff in error variance with increases in intra-frame orientation variance (IFV) of the stimuli. Specifically, errors show a hint of a decrease with IFV for frame 1 but an increase with IFV for frame 2. This overall pattern is a distinct prediction following from subsampling from frame 2



**Fig. 4** Results of subsampling simulations. All data have been mean-centered in order to focus on the pattern of changes with Intra-Frame Variance. The panels show the results from two classes of zero-parameter subsampling models: Random Sampling from all 16 items (top row) and random sampling from only frame 2 (bottom row). Columns correspond to 1-item sampling (column 1), 2-item sampling

(column 2), etc. The results show a distinct advantage for 2-item subsampling. Results for subsampling from frame 1 showed the opposite of the crossover pattern for the frame 2 model, producing systematically poor fits to data. Thus they are omitted here. See text for more details and fit statistics

**Table 5** Differences in subject error variation by interframe interval duration and intraframe variation interactions

IFI (ms)	Frame1xIFI		Frame2xIFI	
	H(3)	p	H(3)	p
30	6.187	0.103	1.458	0.692
38	2.380	0.497	7.756	0.051
46	4.404	0.221	17.452	0.001*
54	3.502	0.320	13.004	0.005*
62	3.499	0.321	17.631	0.001*
70	2.150	0.542	14.940	0.002*
78	10.801	0.013*	16.655	0.001*
86	1.927	0.588	21.578	< 0.001*
	Frame1xFrame2		Frame2xFrame1	
IFV	H(3)	p	H(3)	p
0	5.299	0.150	19.726	< 0.001*
0.25	1.696	0.640	23.543	< 0.001*
0.39	1.576	0.663	5.061	< 0.001*
0.46	6.320	0.100	13.793	< 0.001*

Results that are significant at  $\alpha = 0.05$  are demarcated with an asterisk

The effect in frame 2 also is consistent across the levels of frame 1 IFV (see column denoted *Frame2xFrame1* in the Table). Effects of frame 1 IFV, on the other hand, are weak to nonexistent and do not appear to depend in any systematic way on IFI. For a graphical breakdown of the results shown in Table 5, see Figs. 6 and 7 of the Appendix.

Altogether, it is apparent that, although a full breakdown of the design by all three factors produces much more variable results than the aggregate shown in Fig. 4, they are generally consistent with subsampling of a small number of items from frame 2. Further support comes from inspection of the corresponding, detailed figures in the Appendix, in which it is clear that the bulk of the data show the monotonic, negatively accelerated changes unique to the frame 2 model’s predictions, for both frames with the direction (increase or decrease) determined by the frame. Looking specifically at the shortest IFI, the ambiguity in the empirical results helps to explain the results of the simulations reported earlier, which also showed some ambiguity between random and frame 2 subsampling with 2 items. This appears to be due to temporal integration: participants would not be expected to show a recency effect when they perceive the frames as simultaneous, and would

instead be reading out “randomly” with respect to the frame variable. In short, while precision of mean estimation in this study was independent of temporal integration performance, it does appear that there is at least one “indirect” effect of temporal integration. Specifically, while the mechanisms of integration do not appear to provide a key mechanism of central tendency representation, they do appear to alter the source(s) from which individual items that factor into the average are read into VSTM from iconic memory.

Another possible model is a deterministic selection across frames: selecting 1 item from frame 1 and 1 item from frame 2. We find this unlikely, however, because if participants were systematically carrying out this strategy, we should not have found better performance for frame 2 sampling over frame 1 sampling (or, for that matter, over random sampling, which by definition is a noisy approximation to perfect, deterministic 2-item sampling across both frames).

Another question is whether participants would perform in the same manner if only 2 items were present in the first or second frame. Although we do not empirically test whether a frame with only 2 items would produce the same results as what we observe, the modeling shows that the observer can in principle do this, and we have no reason to suspect they would not give the same results as the models predict. Under our interpretation of the results, the two items sampled are due to the natural limits on information transfer from the iconic to VSTM store, so we predict that displaying only two items would only reduce the ambiguity of which items are sampled into the VSTM store.

In summary, the data are overwhelmingly inconsistent with the idea that participants were subsampling a single item. This means that participants were, in fact, averaging the orientations. However, the modeling suggests that only 2 items were sampled, and they were most likely sampled from frame 2 at all but the two shortest IFIs. These results help to explain why orientation averaging did not depend on IFI as the temporal integration task did. They are also consistent with the fidelity-based integration framework proposed in recent papers (Dubé, 2019; Tong et al., 2019). Specifically, recent work by Tong et al. (2019) has demonstrated that ensemble averaging likely involves a recency- or fidelity-based weighting strategy in which the most recently-viewed 2-3 items carry the vast majority of the weight in the averages provided by participants. Taken together with these prior findings, our results appear to argue against the “preattentive/parallel/global” theory of Treisman and colleagues. Instead, the results appear to support a process in which informational persistence in the iconic store leads to non-selective readout into VSTM, with the averaging operation conducted on the resulting VSTM contents. We elaborate on this in the *General Discussion*, which we turn to next.

## General discussion

We set out to determine whether central tendency representations can be extracted from iconic memory, and whether temporal integration might contribute to the computation. Using a successive-field task with missing and split element reports as well as mean judgments, we found accurate mean estimation that did not vary with inter-frame intervals, over a total stimulus duration ranging from 46ms to 106ms. This held even though the two individual frames containing the to-be-averaged elements had a duration of only 8ms each, followed by a probe mask.

The design, in which only the type of judgment varied across trials, allowed a direct comparison of integration, segregation, and averaging performance. While mean estimation remained steady over all IFIs, integration and segregation performance showed a tradeoff, replicating the findings of Wutz et al. (2016). Taken together, these results seem to suggest that at least some central tendency representations are extracted early, from iconic memory, and that they do not involve classical temporal integration mechanisms. However, a closer inspection of the data suggests a quite different conclusion.

In what follows, we elaborate on the key findings and conclusions from the study, starting with the conclusion that averaging operates on an informational, rather than visible, persistence variable. We note that this opens the door to VSTM, which operates on informational quantities, as the processing locus of averaging. We then discuss evidence from our study and the broader literature that suggests averages are indeed computed later than global pooling and pre-attentive theories assume, using the contents of VSTM rather than iconic memory contents. We conclude by describing a fidelity-based integration account in which feature information held in VSTM is integrated with contributions of individual features to the integration varying based on the fidelity of their VSTM representations.

### Averaging operates on an informational variable

A key finding in our study is that averaging performance was unaffected by IFI, though integration and segregation performance were affected by IFI. The importance of this particular finding is as follows. As detailed in the *Introduction*, iconic memory is believed to comprise two main classes of mechanisms, those underlying visible persistence and those underlying informational persistence. Temporal integration is thought to depend on or measure visible persistence, rather than informational persistence. This idea follows from the finding of disruption of temporal integration when the eyes move during the IFI (Irwin,

1991). Since averaging did not vary with IFI duration in our study, but temporal integration did, we can exclude visible persistence representations as the raw materials of averaging in our task.

This leaves information as the key variable upon which averages are computed. This is an important point as it opens the door to the possibility of a contribution from VSTM, since VSTM depends on output from informational persistence but not from visible persistence (i.e., STM is strictly information-limited, not item- or meaning-limited; Kintsch, 1970).

So, are averages computed using iconic memory contents or VSTM contents? This question is important as the answer will help adjudicate between “early” and “late” accounts of averaging. Having established the informational nature of averaging in our task, we now can approach this theoretically-important question of processing loci. At first glance, the averaging performance in the task is quite good given the extremely short timings used in this iconic memory task. This appears to support early, parallel, pre-attentive, or global pooling accounts of averaging. But on closer inspection, a computation happening later, on VSTM contents, seems more likely. Furthermore, the answer to the question has implications for “subsampling” as well as “early” and “late” classes of theory. We discuss all of this next.

### **Averages are computed “late” and strategic subsampling is just VSTM readout from iconic decay**

At first glance, our data seem to support theories that assume early, global pooling or even pre-attentive extraction of averages. However, we also included a manipulation of the inter-item similarity or “intra-frame variance” within each of the two frames. This allowed us to test distinct (and distinctly differing) predictions of various subsampling models. We found that the variance (1/precision) of participants’ estimation error distributions increased with the IFV of frame 2 and showed some evidence for the opposite pattern, a decrease with IFV, for frame 1. This pattern follows directly from an account in which participants only integrate a few of the most recent items. To determine the number of items being used, we compared the results of 12 models varying in which frame or frames were being used and how many items were being sampled. The results supported an account in which participants sampled 2 items. The results were ambiguous with respect to whether the 2 items were sampled randomly or from frame 2. However, a full breakdown of the empirical data suggested sampling was more likely from the second stimulus array at all but the two shortest IFIs.

What conclusions, then, can be drawn with respect to the mechanisms that are involved in supporting mean estimates? Our results suggest a process by which informational codes representing a small number of the most recently-presented items are non-selectively read out (Averbach & Coriell, 1961) from the iconic store. These codes are transferred to VSTM storage, and it is at this point and with these raw materials that the averaging operation begins. In the current paradigm, we estimate the contents of VSTM as only around 2 items, which may be due to the nature of the rapid presentation of the stimuli. Given the close link between attention and VSTM, it is likely that ensemble processing in this task reflected averaging over attended items that were available in VSTM. As such, our results are inconsistent with the “parallel/preattentive” account and other early accounts of ensemble averaging in which items are globally pooled and averaged prior to attentional selection (Chong & Treisman, 2005b; Whitney & Leib, 2018). Instead, they are broadly consistent with a fidelity-based integration account in which the strongest VSTM memory traces (which will tend to be the most recently-encountered ones) are given priority in the integration or averaging process (Tong et al., 2019).

However, this state of affairs leaves at least one crucial question unanswered. Specifically, are the resulting mean estimates due to control processes or are they due to an implicit mechanism that acts selectively on STM contents? To answer this question, let us first consider the diagnosticity of subsampling behavior for inferring strategic control. That is, subsampling is often referred to as a “strategy” (Myczek & Simons, 2008), but the fact that only a few items contribute to mean estimates does not in and of itself adjudicate between strategic and non-strategic accounts. For instance, consider the present results showing a failure to temporally integrate the items across frames. Even though participants are perfectly able to integrate the arrays (see Fig. 2), they do not appear to use this ability when averaging the stimuli from those same arrays. Instead, the second frame is prioritized. This suggests that what is sampled does not reflect visible persistence mechanisms, but informational persistence mechanisms (Sperling, 1960). In short, what our study suggests is that the limitations on iconic readout and duration of informational traces, which limit the contents of VSTM, are what are producing the “subsample”.

Statistics from information theory provide further evidence in support of this conclusion about the purported “strategic” nature of subsampling. Specifically, estimates of STM encoding rate in adults are typically around  $15 \pm 3$  bits/s (Lehrl & Fischer, 1988; Weiss, 1992). For the simple, 1-dimensional visual stimuli we use in the present work and the total presentation times which range from 46 to 106ms, this amounts to approximately 1 bit of information

on average. So again we see that with the short presentations in use (which force a single readout from iconic memory) we would only expect 2 items to reside in VSTM at the time of responding on each trial, simply from known limitations on STM encoding. Of course, if the participants' mean estimates were extracted early, on iconic storage, we would have expected all of the items to contribute to the participants' responses, as Sperling showed that when iconic memory is directly probed using the partial report procedure, a much larger proportion of items is available than one finds in subsequent VSTM storage from one readout. Hence, the fact that the data show evidence for use of only 2 items is not only an argument against the "strategic" interpretation of subsampling, but it also strengthens the case against a parallel/preattentive account of our results: we find evidence that only 2 items are used, which is the predicted number of items encoded into VSTM. Thus the participants were likely operating on the basis of VSTM contents, not iconic memory contents.

But could it simply be that participants use control processes to modify attention to locations in advance of each array's presentation? We find this exceedingly unlikely. That is, even if we entertain the notion that participants adopted a strategy of ignoring the first frame prior to each presentation, it is hard to imagine volitional attention shifts occurring quickly enough to accommodate the extremely brief IFIs and 8ms frame durations used in this study. Finally, even if participants strategically ignored the first frame, the full set of 8 items in frame 2 would not survive in memory long enough to support VSTM judgments. A "subsample" is precisely what one would expect, without necessitating any strategy at all on the part of the subject, as well as despite any strategy they may try to use. This more or less reduces the explanatory value of strategic control to zero.

In short, subsamples can not by themselves adjudicate between control processes and mechanistic computations. In particular, our study suggests the subsamples are not due to control processes but rather to the normal process of informational readout, which of course results in VSTM for only a subset of the presented items (Averbach & Coriell, 1961; Sperling, 1960). The  $\sim 15$  bits/s encoding rate of STM predicts that if participants use STM contents in our design, we will find only 2 items contribute to their estimates, just as we did. As such, our results question the "strategic" interpretation of subsampling that is frequently stated in the literature (Myczek & Simons, 2008; Simons & Myczek, 2008). Of equal importance, they support the notion that averages are computed "late", upon VSTM contents, rather than upon iconic memory contents. This view is also consistent with the findings of Whiting and Oriet (2011) described in the *Introduction*, in which imposition of a backward mask (which disrupts readout

from iconic memory but not items already in VSTM storage; (Gegenfurtner & Sperling, 1993)) reduced performance in a mean size judgment task. A key prediction of this account, which should be tested in future work, is whether an increase from 2 to 4 items is observed when encoding time in the current paradigm is doubled.

### **An alternative: fidelity-based integration in VSTM**

Having addressed subsampling and the plausibility of "early" averaging theories, we now seek an alternative view that is broadly consistent with our data and prior literature. An analysis of processing stages may be particularly useful in accomplishing this, as it has been to memory theorists in the past (Atkinson & Shiffrin, 1968; Broadbent, 1958). From the point of view of memory processing stages, our results suggest that once the (subsample of) item information surviving the iconic readout and decay process is transferred to VSTM, the averaging process begins. The process itself uses a mechanism of integration by fidelity, which may or may not involve explicit knowledge on the part of the subjects. Specifically, Tong et al. (2019) have reported that item contributions to ensemble averages are prioritized on the basis of their representational fidelity, with more recently-fixated items enjoying a greater contribution to the average (i.e. roughly equivalent to use of a weighted, rather than arithmetic, average; see also Hubert-Wallander and Boynton, 2015). Using a 2-item Sternberg task, Dubé (2019) showed that a mathematical model instantiating a similar assumption was able to outperform the Generalized Context model in accounting for an implicit central tendency effect in those data (Nosofsky et al., 2011). Work using simultaneous tasks, which is ongoing in our lab, can potentially be explained in the same manner, by examining eye movement records and distance of individual items from fixation, and the resulting weight they carry in model-based estimates of integration performance.

In terms of the mechanism and nature of the representation itself, a further consideration of the limitations of integration processes may be informative here. Specifically, as Irwin (1991) pointed out, temporal integration is hindered when there is a change in retinal position across the successive fields. For this reason, Irwin argued that temporal integration is not responsible for perceived visual continuity across fixations. Instead, Irwin's data suggested an abstract informational form of integration, likely involving the construction of an "object file" representation of the sort described subsequently by Kahneman et al. (1992).

The lack of dependence on integration mechanisms in the current study suggests that, of the known iconic memory subcomponents, it is the informational, rather than visible,

persistence component that provides the raw materials for the mean computation. If this is true, then a more detailed model of our results immediately follows: informational persistence provides abstract codes to VSTM. These codes are then subject to a central, non-retinotopic integration mechanism which creates an abstract representation akin to an “object file”, with items factoring into the computation according to the fidelity of their memory representations. The resulting representation is either the central tendency representation itself or may be a concatenation of feature, context, item, and statistical information.

Assuming, then, that we pursue this fidelity-based integration view, we must ask: how does it fare with respect to the “parallel and preattentive” and global pooling theories that are so frequently encountered in the ensemble coding literature?

Fidelity-based integration has the advantage of consistency with the broader literature on ensemble coding. Specifically, it is clear that participants can produce accurate mean estimates in tasks for which iconic memory is unlikely to contribute. More specifically, several studies of mean estimation have included sequential presentation of the to-be-averaged items using long stimulus durations and ISIs on the order of seconds (Tong et al., 2019). Some of these investigations used extremely long lags between the to-be-averaged items and the probe, in tasks in which participants were unaware, prior to the probe, that they would be asked to generate a mean estimate (Oriet & Hozempa, 2016). The accurate performance of the judgments in these tasks effectively rules out early, automatic iconic integration as the only mechanism involved.

Putting this all together, we see that if we accept the parallel/preattentive model, we are confronted with two diametrically opposed lines of evidence regarding ensemble coding mechanisms. On one hand we have the evidence described in the *Introduction* supporting early, automatic extraction from iconic memory without downstream contributions, and on the other hand we have the evidence supporting the contribution of control processes acting on visual short-term memory contents without a likely contribution from early, iconic memory processes. This suggests that either a single, supramodal mechanism makes contact at multiple points along the processing stream, or that the various mean judgment tasks are each tapping into one of two different processes (control processes acting on VSTM contents, and rapid iconic extraction processes) and only carry a superficial relation in the form of a shared overt response requirement.

Further difficulties arise if we include within our scope the large literature, originating before modern studies of “ensemble coding”, that shows an implicit bias from ensemble means in judgments that are nominally unrelated to the statistics of prior or concurrent stimulus features

(Hollingworth, 1910; Dubé, 2019, 2015). One would then be compelled to add to the collection of mechanisms an implicit learning mechanism that integrates information into a central tendency trace. This is the price to be paid for retaining the automatic/preattentive account: we must invent a new mechanism for every kind of ensemble coding task (simultaneous, sequential, iconic, long-term, implicit, explicit, etc), even though those tasks produce very similar kinds of responses (mean judgments and item judgments biased by the mean).

If we instead accept the view suggested by the current study, it appears possible that all of the findings can be subsumed under one fidelity-based integration mechanism (Tong et al., 2019) operating on STM to create an abstract, central trace for storage in LTM. This is the advantage of abandoning the automatic/preattentive account.

Our view is of course not the only conceivable alternative. Recent work by Baek and Chong (2020a, b) has advanced an elegant distributed attention account that brings aspects of the “pre-attentive/automatic” model of Chong and Treisman (2003) into precise quantitative focus. Using the well-known engineering framework of Lu and Doshier (1998), the authors instantiated several models varying in the loci of internal noise and the contributions of two separate models of attentional selection. Model selection was then conducted using new data with the 2-frame mean comparison task as well as some prior data. Of these various models, a “Zoom Lens” model provided the best account of the data. For instance, the model correctly predicted improved performance with increases in set size.<sup>3</sup> Since this model allows attention to spread dynamically over a surface, the authors concluded that all items contribute equally to mean computation.

Although Baek and Chong’s model in our view represents a significant advance over prior work, it does not provide us with a clear picture from the perspective of processing stages. Specifically, stages of encoding and transfer from the sensory store to short-term memory are not specified in the model. For this reason, though the model is helpful in answering some of the important questions posed by Baek and Chong, it is less helpful in answering the questions we are asking in the current work. Specifically, we are asking “what memory system or systems are involved in averaging.” Since the Zoom Lens model does not clearly specify an iconic memory system or quantify the read-out process to VSTM, the Zoom Lens model in its current form

<sup>3</sup>We suspect that the longer presentation times used in the cited studies (e.g. Baek and Chong (2020a), which used around 500ms for each of two arrays) may be an important factor. These timings allow greater readout to VSTM because multiple passes through iconic storage are possible (Averbach & Coriell, 1961; Sperling, 1960). Such designs are therefore ill-suited to address our current goal: to determine whether averages can be extracted prior to VSTM storage.

does not appear equipped to answer the questions we are asking.

Work using the centroid (center of mass) estimation task provides additional insights relevant to our conclusions. Specifically, a series of papers by Chubb, Sperling and colleagues (Drew et al., 2010; Inverso et al., 2016; Rodriguez-Cintron et al., 2019; Sun et al., 2016a, b, 2018, e.g.) has investigated the attentional filtering operations used by participants to locate the centroid of clouds of briefly-presented dots, with the dots typically varying in some target feature such as color. Of particular interest in this series is the recent paper by Sun et al. (2018) in which the authors provide evidence that centroids may be computed “preconsciously”. Most of the experiments involved a 300ms display of 26 colored dots, in which roughly half of the dots shared the same color (e.g. red dots) and the remaining dots were heterogeneously-colored (a green dot, a blue dot, etc). Participants were asked, in separate blocks of trials, to indicate via mouse-click the centroid location of either the homogenous or heterogeneous dots (Exp. 1), both centroids, one click at a time (Exp. 2), or the centroid over all dots (Exp. 3). Importantly, the dots were randomly, spatially intermingled on each trial, forcing participants to rapidly segregate the items by a group feature (similarity or dissimilarity) prior to computing the centroid.

The results were striking. The authors found that participants were highly efficient in all of these tasks. Estimates of efficiency were obtained in a manner similar to the subsampling simulations we report, and produced lower bounds on the number of items necessary to reach human performance. These efficiencies fell in a range of lower bounds from 8 to 18 items contributing to participants’ performance, depending on the task. Analysis of attention operating characteristics showed further that the requirement in Exp. 3 to carry out two centroid computations on the same intermingled dot array produced little to no drop in performance or efficiency estimates. Post-exposure masking also had little to no effect on the results. Finally, in a fourth experiment, the authors measured change detection ability for the dot stimuli by presenting the arrays twice: Once for 300ms, followed by a pattern mask, and then again as a probe stimulus. The probe stimulus, however, had two dots either missing or added, relative to the first presentation. Change detection accuracy was reported to be “shockingly poor” and estimates of efficiency suggested a single dot, at best, contributed to performance.

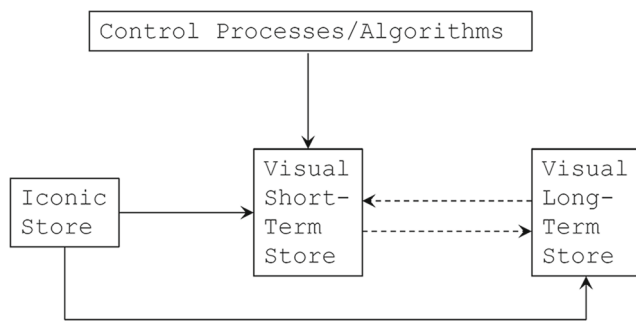
The results of Sun et al. (2018) do strongly suggest that centroids of spatial arrays are computed very rapidly, perhaps “preconsciously” as the authors suggest. This indicates that our own conclusions from the present work may not extend beyond the orientation stimuli we have used, which is a clear limitation of the current work.

However, the existing evidence suggests that the boundaries of generalizability may be more tightly circumscribing the centroid task, rather than the orientation task. For instance, other work by Chubb, Sperling and colleagues has shown that size averaging shows much lower efficiency than centroid estimation (Rodriguez-Cintron et al., 2019), and that centroid estimation performance degenerates if a feature other than color (i.e. orientation, which we use in our work) is used as the target-defining feature (Inverso et al., 2016). Finally, it is worth considering the relation of the centroid task to classic partial report tasks. The latter have shown that when a letter row or column is cued either in advance or concurrently with the display onset, attentional selection can occur (“selective” as opposed to “non-selective” readout; Averbach & Coriell, 1961). In other words, attention can filter iconic readout to VSTM. From this standpoint, the centroid task used by Sun et al. (2018) may be viewed as a kind of partial-report task with a concurrent color homogeneity cue, with the task being essentially to read out the central row and column of an array. From this point of view, the results may not be as surprising as they at first appear. To be certain that the task has anything to do with central tendency representation, for instance, it would first be necessary to directly compare the centroid task with reporting of the location of other “quadrants” of the array. Nonetheless, the results of the centroid studies do suggest important limitations to our conclusions.

## Conclusions

We have found that orientation averaging does not appear to depend on temporal integration mechanisms, suggesting informational persistence provides the raw materials of averaging. Furthermore, simulations and information theory statistics, along with known properties of VSTM (e.g. its information-, rather than item- or meaning-, limited capacity) suggest that our results reflect an averaging operation conducted on VSTM contents, rather than pre-attentive, global pooling, or other “early” mechanisms. We describe a fidelity-based integration theory that appears to simplify explanation of our findings with respect to the broader literature on central tendency estimation. We note that accepting a pre-attentive or early view entails a less parsimonious explanation of these various findings. The current state of affairs is diagrammed using the modal model (Atkinson & Shiffrin, 1968) in Fig. 5, with the likely routes by which computations may be transferred highlighted. Specifically, all routes are necessary once an early, preattentive, or global pooling account is adopted; only the dotted routes are necessitated under a fidelity-based account. We hope that work on ensemble coding, which has yet to provide a comprehensive mathematical theory,





**Fig. 5** The Atkinson-Shiffrin (1968) model. Potential loci of ensemble code transfer include two routes from iconic storage. Ensemble codes may also be created via control processes acting on the contents of visual short-term memory. The existing literature provides support for some form of computation acting upon or within the iconic store alone, as well as for control processes acting on visual short-term memory contents without any direct involvement of the iconic store, as well as for implicit learning and LTM storage and retrieval of ensemble codes. We suggest a single, fidelity-based integration mechanism operating on VSTM contents is simpler. Possible transfer routes for the resulting codes in this account are shown as dashed lines

will produce further novel insights into the structural loci and functional forms underlying the creation, maintenance, and retrieval of ensemble representations such as mean representation, which we have focused on here.

**Open Practices Statement** The data and materials for the experiments reported here are available upon request. None of the experiments were preregistered.

### Appendix: Stimulus distributions

The set of angles that comprised each stimulus frame was defined by one of 16 unique circular mean and variance combinations (see Table 1). The mean values for the frames with  $\sigma_c^2 = 0$ , by definition, took the mean value of the singular shared individual stimulus orientation ( $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , or  $135^\circ$ ). The mean values that make up each of the 3 non-zero variance level combinations were kept as approximately close to one another as possible, such that they may be considered categorically grouped into 4 levels of comparable mean values.

As there were four unique approximations of  $\mu_c$  for  $\sigma_c^2 > 0$  this amounts to  $\binom{4}{2} = 6$  plus 4 self-pairings for a total of 10 unique combinations of means (ignoring order) across frames 1 and 2, for  $\sigma_c^2 > 0$ . To this are added the 10 combinations ( $\binom{4}{2} = 6$ , plus 4 self-pairings) for angle sets with  $\sigma_c^2 = 0$ , and 16 more for combinations between the  $\sigma_c^2 = 0$  and  $\sigma_c^2 > 0$  classes of angle sets. The combination of means on each trial was thus a random draw from this full set of 36 unique combinations. However, the probabilities are biased in favor of the  $\sigma_c^2 > 0$  class of combinations, with the odds being  $\frac{9}{1}$  in favor of seeing the  $\sigma_c^2 > 0$  combination

class rather than the  $\sigma_c^2 = 0$  combination class, and  $\frac{3}{1}$  against the cross-set combination class.

Finally, as noted in *Procedure*, there was a total of 240 mean estimation trials, and 4 non-contiguous repetitions of that task block in the experiment. Thus, within a given mean estimation block, if each mean combination were represented equally often there would be 7 presentations of the entire class of 36 unique combinations. However, because of the repeated set of four approximations to the mean across the four nonzero variance conditions, the total number of combinations (including non-unique, repeated combinations), is 148 (10 for the zero class, 3 presentations of 16 for the cross-class, and 9 presentations of 10 for the  $> 0$  class).

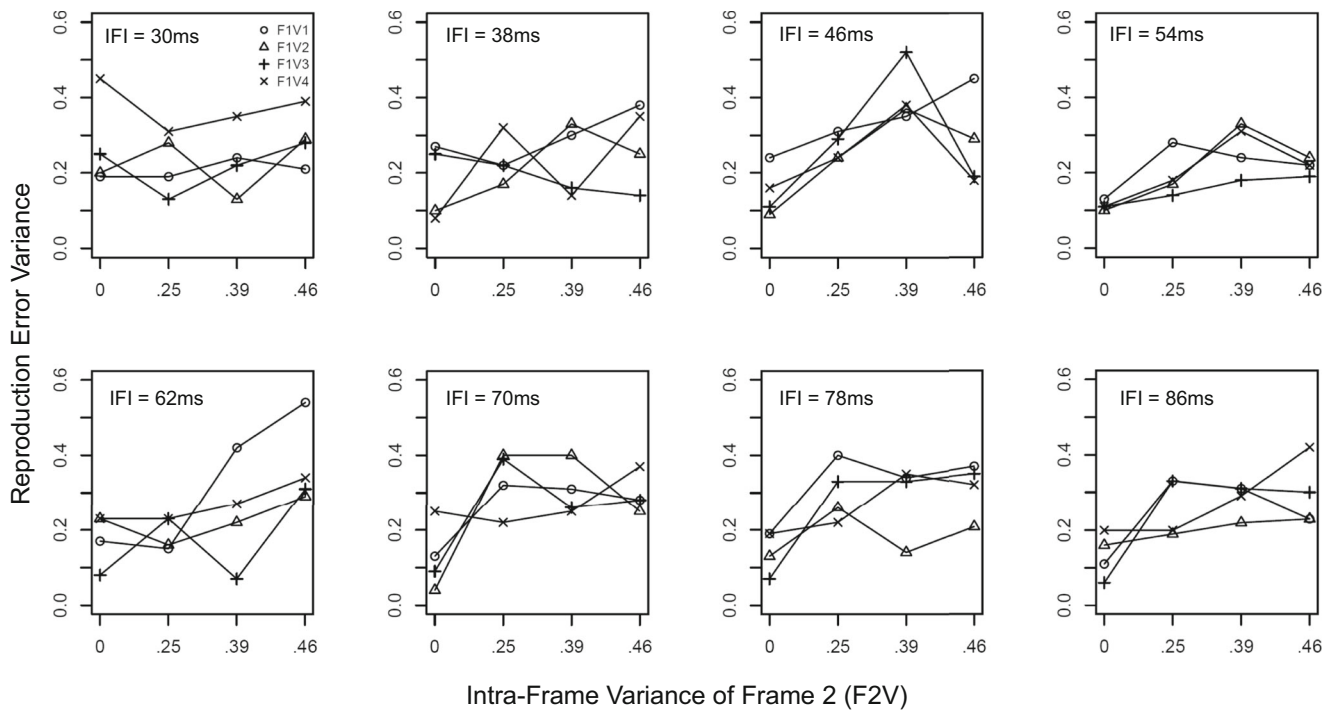
### Intra-frame variance, intra-frame interval, and frame number

We investigated 12 subsampling models, with zero free parameters (as described in *Subsampling*). The models differed in the number of items that were sampled, crossed with the sampling pool (all 16 items, frame 1 only, or frame 2 only). We simulated datasets with all possible set sizes, but found consistently poor performance across all models for set sizes beyond 4 items subsampled.

As shown earlier in Fig. 4, the 2-item frame 2 model provided as good a fit as the random 2-item model. However, the frame 2 model appeared to provide a better description of the form of the data. That is, the data appeared to show a negatively-accelerated increase in error variance as a function of frame 2 IFV, but when plotted against frame 1 IFV the model showed a negatively-accelerated decrease in error variance while the data suggested a non-monotonic “U” shape.

In this section, we provide a more detailed look at the empirical data, which suggest that the frame 2 model was making the correct prediction for the bulk of the frame 1 data.

Graphical depictions of the interaction between the interframe interval (IFI), and frame 1 and frame 2 intra-frame variance (IFV) are included below. The figures display a great deal of apparent variability in the data; however, it does appear that reproduction error variance generally increases with frame 2 IFV and this trend does not appear to depend systematically on IFI, with the exception of the shortest IFI of 30ms. Here, there is an increase in error variance when frame 1 has a high IFV, but no consistent pattern for Frame 2 with IFV. As described previously, the most consistent effect is the increase in error variance with increases in IFV of frame 2, with a recency effect appearing based on the temporal segregation of the two frames at post-threshold IFI.



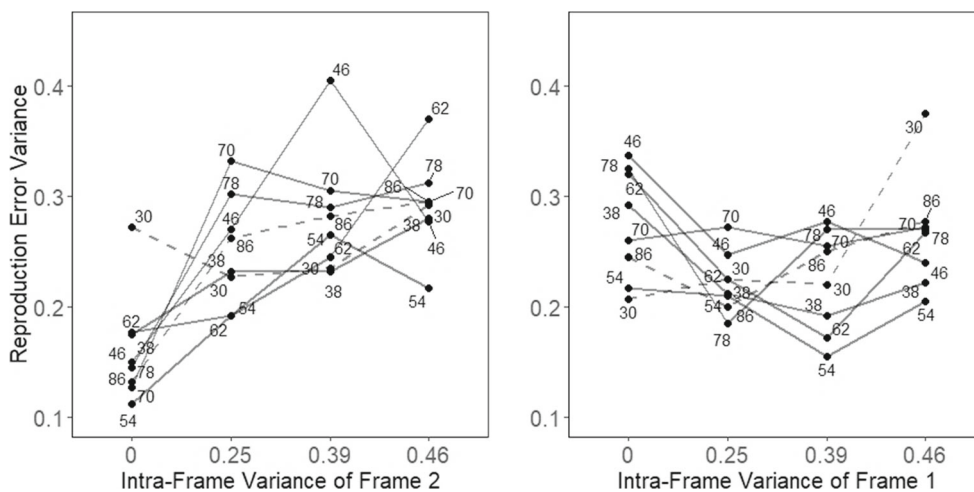
**Fig. 6** Observed reproduction error variance as a function of frame 1 and 2 intra-frame variance (IFV) and duration of the interval between frames (IFI). The labels “F1VX” refer to frame 1 IFV levels, where X denotes the level (1 for  $\sigma_c^2 = 0$ , 2 for  $\sigma_c^2 = .25$ , etc.). The results show an increase in error variance with the intra-frame variance of frame 2 which does not depend systematically on inter-frame interval. The patterns in error variance with increases in frame 1 intra-frame variance

are less stable or clear. For the IFI = 30ms condition, however, there is no obvious pattern for frame 2 although a trend is apparent showing higher error at the highest level of intra-frame variance of frame 1. Overall, the results are consistent with the those reported in Fig. 4, which collapsed over conditions to reduce noise in the data, although they do suggest that calculation based on readout from frame 1 may be slightly more likely at the shortest IFI

The trends are broadly consistent with the results of the simulations from the 2-item frame 2 model. Specifically, it is clear that the one departure from the model, the apparent “U” shape across frame 1 IFV (see Fig. 4), is the result of a single large departure from the general trend, which is the

increase in error for the 30ms IFI condition when frame 1 has a large IFV (IFV = .46).

In short, closer inspection of the data reveals even stronger support for the 2-item, frame 2 model than was apparent in the earlier analysis.



**Fig. 7** The data from Fig. 6, collapsed over frame 1 (left panel) or frame 2 (right panel). Shortest and longest IFIs are shown with dotted lines. The data again show that response error increases with frame 2 IFV and shows a slight decrease for frame 1 IFV. The data do not show a clear pattern at the shortest IFI, however. Across the longer IFIs, there is no apparent change in the observed increases or decreases. In other words, these results, which aggregate over frame 1 or 2, are nonetheless consistent with those of Fig. 6

## References

- Albrecht, A. R., & Scholl, B. J. (2010). Perceptually averaging in a continuous visual world: Extracting statistical summary representations over time. *Psychological Science*, *21*(4), 560–567.
- Alvarez, G. A., & Oliva, A. (2007). The representation of ensemble visual features outside the focus of attention. *Journal of Vision*, *7*(9), 129–129.
- Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398.
- Alvarez, G. A., & Oliva, A. (2009). Spatial ensemble statistics are efficient codes that can be represented with reduced attention. *Proceedings of the National Academy of Sciences*, *106*(18), 7345–7350.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties, *12*(2), 157–162. Retrieved from <http://pss.sagepub.com/content/12/2/157>.
- Atkinson, R., & Shiffrin, R. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Elsevier.
- Attarha, M., Moore, C. M., & Vecera, S. P. (2016). The time-limited visual statistician. *Journal of experimental psychology: Human Perception and Performance*, *42*(10), 1497.
- Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *The Bell System Technical Journal*, *40*(1), 309–328.
- Baek, J., & Chong, S. C. (2020a). Distributed attention model of perceptual averaging. *Attention, Perception, Psychophysics*, *82*(1), 63–79.
- Baek, J., & Chong, S. C. (2020b). Ensemble perception and focused attention: Two different modes of visual processing to cope with limited capacity. *Psychonomic Bulletin Review*, 1–5.
- Baijal, S., Nakatani, C., Leeuwen, C., & van Srinivasan, N. (2013). Processing statistics: An examination of focused and distributed attention using event related potentials. *Vision Research*, *85*, 20–25.
- Beepa (2013). FRAPS [Computer Software] Version 3.5.99. Beepa Pty Ltd. Retrieved from <https://fraps.com/>.
- Bondarko, V. M., & Danilova, M. V. (1997). What spatial frequency do we use to detect the orientation of a landolt c? *Vision Research*, *37*(15), 2153–2156.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial vision*, *10*(4), 433–436.
- Broadbent, D. (1958). *Perception and communication*. Elmsford: Pergamon Press.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.
- Chong, S. C., & Treisman, A. (2005a). Attentional spread in the statistical processing of visual displays. *Perception Psychophysics*, *67*(1), 1–13.
- Chong, S. C., & Treisman, A. (2005b). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*(7), 891–900.
- Chong, S. C., Joo, S. J., Emmanouil, T.-A., & Treisman, A. (2008). Statistical processing: Not so implausible after all. *Perception Psychophysics*, *70*(7), 1327–1334.
- Coltheart, M. (1980). Iconic memory and visible persistence. *Perception Psychophysics*, *27*(3), 183–228.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity, *24*(1), 87–114; discussion 114–185.
- De Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception Psychophysics*, *70*(5), 789–794.
- Dick, A. (1969). Relations between the sensory register and short-term storage in tachistoscopic recognition. *Journal of Experimental Psychology*, *82*(2), 279.
- Drew, S. A., Chubb, C. F., & Sperling, G. (2010). Precise attention filters for weber contrast derived from centroid estimations. *Journal of vision*, *10*(10), 20–20.
- Dubé, C., & Sekuler, R. (2015). Obligatory and adaptive averaging in visual short-term memory. *Journal of Vision*, *15*(4), 13–13.
- Dubé, C. (2019). Central tendency representation and exemplar matching in visual short-term memory. *Memory Cognition*, *47*(4), 589–602.
- Efron, R. (1970). The minimum duration of a perception. *Neuropsychologia*, *8*(1), 57–63.
- Emmanouil, T. A., & Treisman, A. (2008). Dividing attention across feature dimensions in statistical processing of perceptual groups. *Perception Psychophysics*, *70*(6), 946–954.
- Eriksen, C. W., & Collins, J. F. (1967). Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology*, *74*(4p1), 476–484.
- Fisher, N. (1993). *Statistical analysis of circular data*. Cambridge University Press.
- Gegenfurtner, K. R., & Sperling, G. (1993). Information transfer in iconic memory experiments. *Journal of Experimental Psychology: Human Perception and Performance*, *19*(4), 845.
- Haber, R. N., & Standing, L. (1969). Direct measures of short-term visual storage. *Quarterly Journal of Experimental Psychology*, *21*(1), 43–54.
- Harrison, W. J., & Bex, P. J. (2015). A unifying model of orientation crowding in peripheral vision. *Current Biology*, *25*(24), 3213–3219.
- Hogben, J. H., & Lollo, V. di. (1974). Perceptual integration and perceptual segregation of brief visual stimuli. *Vision Research*, *14*(11), 1059–1069.
- Hollingworth, H. L. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, *7*(17), 461–469.
- Hubert-Wallander, B., & Boynton, G. M. (2015). Not all summary statistics are made equal: Evidence from extracting summaries across time. *Journal of vision*, *15*(4), 5–5.
- Inverso, M., Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2016). Evidence against global attention filters selective for absolute bar-orientation in human vision. *Attention, Perception, Psychophysics*, *78*(1), 293–308.
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*(3), 420–456.
- Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive Psychology*, *24*(2), 175–219.
- Kintsch, W. (1970). *Learning, memory, and conceptual processes*. Wiley.
- Kleiner, M., Brainard, D. H., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*.
- Lehrl, S., & Fischer, B. (1988). The basic parameters of human information processing: Their role in the determination of intelligence. *Personality and Individual Differences*, *9*(5), 883–896.
- Long, G. M. (1980). Iconic memory: A review and critique of the study of short-term visual storage. *Psychological Bulletin*, *88*(3), 785–820.
- Lu, Z.-L., & Doshier, B. A. (1998). External noise distinguishes attention mechanisms. *Vision research*, *38*(9), 1183–1198.
- 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 psychologica*, *142*(2), 245–250.

- McAnany, J. J., & Alexander, K. R. (2008). Spatial frequencies used in landolt c orientation judgments: Relation to inferred magnocellular and parvocellular pathways. *Vision Research*, *48*(26), 2615–2624.
- Mewhort, D., Campbell, A., Marchetti, F. M., & Campbell, J. I. (1981). Identification, localization, and “iconic memory”: An evaluation of the bar–probe task. *Memory & Cognition*, *9*(1), 50–67.
- Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788.
- Nosofsky, R. M., Little, D. R., Donkin, C., & Fific, M. (2011). Short-term memory scanning viewed as exemplar–based categorization. *Psychological review*, *118*(2), 280.
- Oriet, C., & Brand, J. (2013). Size averaging of irrelevant stimuli cannot be prevented. *Vision Research*, *79*, 8–16.
- Oriet, C., & Hozempa, K. (2016). Incidental statistical summary representation over time. *Journal of Vision*, *16*(3), 3–3.
- Pelli, D. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Rodriguez-Cintrón, L. M., Wright, C. E., Chubb, C., & Sperling, G. (2019). How can observers use perceived size? centroid versus mean–size judgments. *Journal of vision*, *19*(3), 3–3.
- Simons, D. J., & Myczek, K. (2008). Average size perception and the allure of a new mechanism. *Perception & Psychophysics*, *70*(7), 1335–1336.
- Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29.
- Sperling, G. (1967). Successive approximations to a model for short term memory. *Acta Psychologica*, *27*, 285–292.
- Stroud, J. M. (1956). The fine structure of psychological time. In Quastler, H. (Ed.), *Information theory in psychology: Problems and methods*. Glencoe, Ill.: Free Press.
- Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2016a). The centroid paradigm: Quantifying feature–based attention in terms of attention filters. *Attention, Perception, & Psychophysics*, *78*(2), 474–515.
- Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2016b). Human attention filters for single colors. *Proceedings of the National Academy of Sciences*, *113*(43), E6712–E6720.
- Sun, P., Chubb, C., Wright, C. E., & Sperling, G. (2018). High–capacity preconscious processing in concurrent groupings of colored dots. *Proceedings of the National Academy of Sciences*, *115*(52), E12153–E12162.
- Tiurina, N., & Utochkin, I. (2014). Visual size averaging is parallel but depends on the range. *Journal of Vision*, *14*(10), 876–876.
- Tong, K., Dubé, C., & Sekuler, R. (2019). What makes a prototype a prototype? averaging visual features in a sequence. *Attention, Perception, & Psychophysics*, *81*(6), 1962–1978.
- Treisman, A. (1985). Preattentive processing in vision. *Computer vision, graphics, and image processing*, *31*(2), 156–177.
- Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological review*, *95*(1), 15.
- Treisman, A., Vieira, A., & Hayes, A. (1992). Automaticity and preattentive processing. *The American journal of psychology*, *341*–362.
- Weiss, V. (1992). The relationship between short–term memory capacity and eeg power spectral density. *Biological Cybernetics*, *68*(2), 165–172.
- Weyer, E. (1899). Die zeitschwellen gleichartiger und disparater sinneseindrücke. [temporal thresholds for intra–and inter-modal sensory impressions]. *Philosophische Studien*, *15*, 68–138.
- White, C. T. (1963). Temporal numerosity and the psychological unit of duration. *Psychological Monographs: General and Applied*, *77*(12), 1–37.
- Whiting, B. F., & Oriet, C. (2011). Rapid averaging? not so fast! *Psychonomic Bulletin & Review*, *18*(3), 484–489.
- Whitney, D., & Leib, A. Y. (2018). Ensemble perception, *69*(1), 105–129.
- Wundt, W. (1900). Zur kritik tachistoskopischer versuche. *Philosophische Studien*, *16*, 61–70.
- Wutz, A., Muschter, E., van Koningsbruggen, M. G., Weisz, N., & Melcher, D. (2016). Temporal integration windows in neural processing and perception aligned to saccadic eye movements. *Current Biology*, *26*(13), 1659–1668.
- Zar, J. (1999). *Biostatistical analysis*, (4th ed.). Prentice Hall.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.