

Perceiving set mean and range: Automaticity and precision

Noam Khayat

Life Sciences Institute and
Edmond and Lily Safra Center (ELSC) for Brain Research,
Hebrew University, Jerusalem, Israel



Shaul Hochstein

Life Sciences Institute and
Edmond and Lily Safra Center (ELSC) for Brain Research,
Hebrew University, Jerusalem, Israel



To compensate for the limited visual information that can be perceived and remembered at any given moment, many aspects of the visual world are represented as summary statistics. We acquire ensemble representations of element groups as a whole, spreading attention over objects, for which we encode no detailed information. Previous studies found that different features of items (from size/orientation to facial expression/biological motion) are summarized to their mean, over space or time. Summarizing is economical, saving time and energy when the environment is too rich and complex to encode each stimulus separately. We investigated set perception using rapid serial visual presentation sequences. Following each sequence, participants viewed two stimuli, member and nonmember, indicating the member. Sometimes, unbeknownst to participants, one stimulus was the set mean, and or the nonmember was outside the set range. Participants preferentially chose stimuli at/near the mean, a “mean effect,” and more easily rejected out-of-range stimuli, a “range effect.” Performance improved with member proximity to the mean and nonmember distance from set mean and edge, though they were instructed only to remember presented stimuli. We conclude that participants automatically encode both mean and range boundaries of stimulus sets, avoiding capacity limits and speeding perceptual decisions.

Introduction

Interacting with the environment, our visual system is constantly confronted with a large, dynamic stream of information that exceeds its processing capacity (Ariely, 2001; Cohen, Dennett, & Kanwisher, 2016; Robitaille & Harris, 2011). Unlike situations in which we focus visual attention on one or a few objects, as limited by the cognitive mechanisms of attention and

working memory (Cowan, 2001; Luck & Vogel, 1997), usually we need to account for many objects containing multiple properties. In particular, sometimes we spread attention globally over a set of items with some degree of similarity—a crowd of animals at the safari, a shelf of alcohol bottles at a bar, a line of cars in traffic, or a copse of trees in a forest. After looking away from the scene, we have only limited conscious access to information about individuals we just saw (Fabre-Thorpe, 2011). Rather, we have a global representation of the set as a whole (Alvarez & Oliva, 2008; Ariely, 2001; Corbett & Oriet, 2011). The visual system maximizes the limited attentional resources available for processing salient objects or events (Cohen, Dennett, & Kanwisher, 2016; Jackson-Nielsen, Cohen, & Pitts, 2017), while maintaining the “gist” of the surrounding environment, the essence of the scene (Alvarez & Oliva, 2009; Hochstein & Ahissar, 2002). This is accomplished, in part, by rapidly extracting sets of items, spatially and temporally, forming summarized representations (average, range, or variance) of their features. Processing statistical properties helps form a coherent representation of the global visual scene, circumventing the limiting bottleneck in coding more than about four objects for detailed conscious analysis (Cowan, 2001; Luck & Vogel, 1997).

Summary statistics perception appears to be a general mechanism operating on various stimulus attributes, and may represent a different mode of visual processing affording observers an ensemble percept and access to information about the gist of the entire scene, enabling them to perceive far more than a few objects at a time, overcoming capacity limitations (Ariely, 2001; Cohen et al., 2016). Extraction occurs quickly and perhaps automatically (i.e., without being a conscious goal of the performed task), and has been shown to be easier to combine with tasks requiring

Citation: Khayat, N., & Hochstein, S. (2018). Perceiving set mean and range: Automaticity and precision. *Journal of Vision*, 18(9):23, 1–14, <https://doi.org/10.1167/18.9.23>.

<https://doi.org/10.1167/18.9.23>

Received April 8, 2018; published September XX, 2018

ISSN 1534-7362 Copyright 2018 The Authors



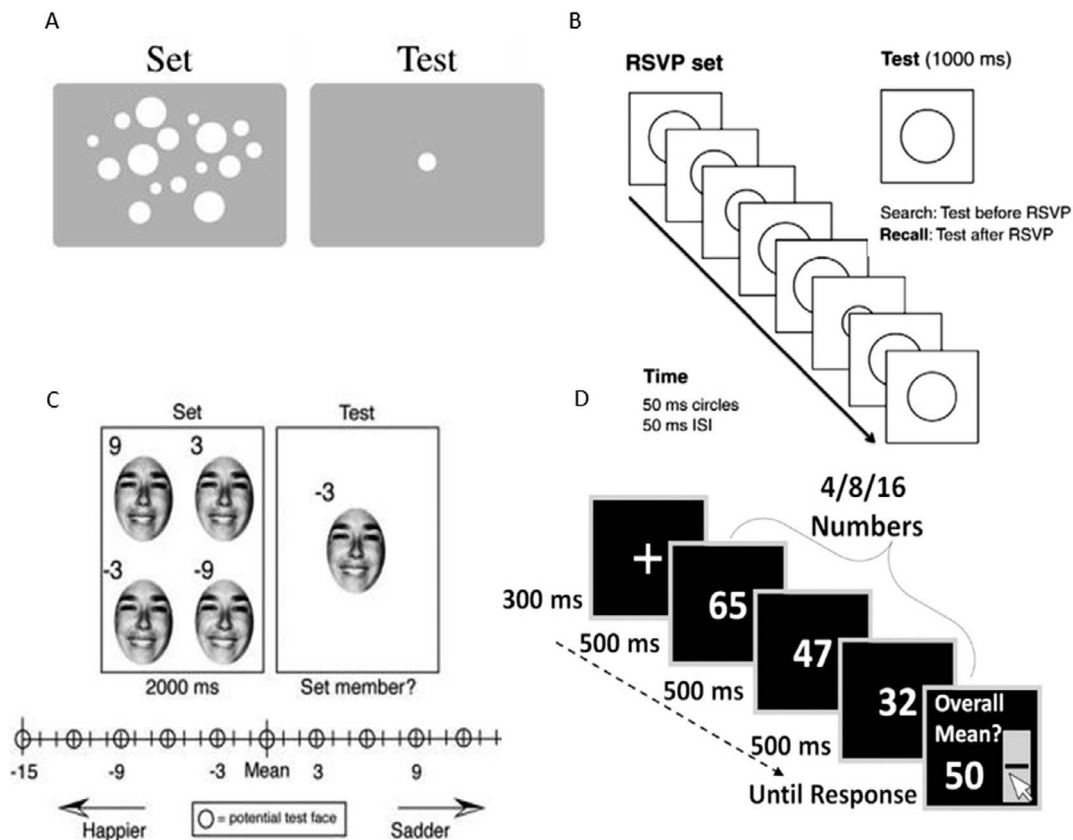


Figure 1. Testing set summary perception. (A) Two trial intervals, presenting a set of circles followed by a test circle asking if the test was present or if it equals the set mean (Ariely, 2001). (B) RSVP sequence of 5–11 circles, with a test circle before or after the sequence (Corbett & Oriet, 2011). (C) Faces (4–16) varying in emotional expression, followed by a test face (Haberman & Whitney, 2009). (D) A set of (4–16) two-digit numbers sequentially presented, asking participants to report set average (Brezis et al., 2015).

distributed rather than focused attention (Alvarez & Oliva, 2009; Chong & Treisman, 2005).

Visual information statistically represented

Summary statistics (also referred to as ensemble coding or set representations) have been discussed generally in the context of experiments displaying simple stimuli as circles (Ariely, 2001; Corbett & Oriet, 2011; see Figure 1A, 1B), Gabor patches (Attarha & Moore, 2015a) and colored forms (Ward, Bear, & Scholl, 2016). These tested, correspondingly, coded information of mean size (Ariely, 2001; Corbett & Oriet, 2011; Alik, Toom, Raidvee, Averin, & Kreegi-puu, 2014), orientation (Alvarez & Oliva, 2009), and hue (Maule & Franklin, 2015; Ward et al., 2016), as well as brightness (Bauer, 2009), spatial position (Alvarez & Oliva, 2008), and motion speed and direction (Sweeny, Haroz, & Whitney, 2013). Most studies used static arrays (Ariely, 2001; Chong & Treisman, 2003, 2005), in which set stimuli were presented simultaneously, testing spatial summation (Alvarez & Oliva, 2009). Nevertheless, visual input in

real life is dynamic (Hubert-Wallander & Boynton, 2015), and other studies used rapid visual serial presentation (RSVPs), requiring summarizing over time (Corbett & Oriet, 2011; Brezis, Bronfman, & Usher, 2015; Hubert-Wallander & Boynton, 2015). With either presentation mode, results show that observers estimate mean set properties quite precisely and performance is speeded with increased number of display items (Corbett & Oriet, 2011; Robitaille & Harris, 2011). In contrast, observer ability to identify set member items was close to chance, indicating that they coded very little information of individuals but still possessed information concerning the set as a whole (Ariely, 2001; Corbett & Oriet, 2011; Utochkin, 2015; Ward et al., 2016).

Extraction of summary statistics is not exclusive for perception of low-level stimuli. Set mean perception occurs also for high-level information, such as facial emotion (Haberman & Whitney, 2007, 2009; Figure 1C; Neumann, Schweinberger, & Burton, 2013), object lifelikeness (Yamanashi Leib, Kosovicheva, & Whitney, 2016) and biological motion of human crowds (Sweeny, Haroz, & Whitney, 2013), suggesting that summary representations are generated in high-level

visual areas as well. This ability is also relevant to auditory stimuli, summarizing pitch or frequency sequences (Albrecht, Scholl, & Chun, 2012; Piazza, Sweeny, Wessel, Silver, & Whitney, 2013; McDermott, Schemitsch, & Simoncelli, 2013), and we are also able to estimate numerical averages of two-digit numbers (Brezis et al., 2015; see Figure 1D), using a remarkable approximate holistic “intuitive mode.” Taken together, these suggest that averaging might be a general perceptual ability.

Experimental goals

We first confirm and expand conclusions of the above studies, testing the impact of mean presence in a membership task, with two essential innovations: We test memory with a two-alternative forced choice (2-AFC) paradigm, allowing a criterion-free result; (every correct response is both a hit and a correct rejection; every incorrect response both a miss and a false alarm). Furthermore, instead of executing both membership and mean tests, in the current experiment we instructed participants to perform only membership tests, as if examining their visual memory, so that perception of the mean is implicit. On each experimental trial, participants viewed a centrally-presented RSVP sequence of low-level elements with a varying feature. Then, they were shown two test elements and asked to select which was a member of the sequence. By placing the mean set element as one of the test elements (member or nonmember), we tested how the mean property (or proximity to the mean) influences the results. Participants were not informed of the involvement of mean properties and the division to different trial subtypes. Combining this strategy and short RSVP exposure times, we limit observer ability of processing individual elements, leaving only gist perception of the set. We hypothesize that as the member test element is closer to the mean set property, performance will improve, and as the nonmember test element is closer to the mean, performance will deteriorate. This trend is termed the “Mean Effect.”

The second, less discussed property of summary statistics is perception of set range or its boundaries. It has been shown that observers perceive the variance of stimulus sets (Dakin & Watt, 1997; Morgan, Chubb, & Solomon, 2008; Solomon, 2010), and there is indirect evidence concerning observer knowledge of the set range, itself (Ariely, 2001; Haberman & Whitney, 2010), and its effect on mean judgment (Maule & Franklin, 2015). Here, we directly test range-perception effects on membership judgment, termed the “Range Effect.” To observe this effect, the nonmember test element was occasionally an element that is outside of the set range. For example, in circle size trials, the

nonmember element could be either larger than the largest set circle or smaller than the smallest one.

We performed three psychophysical experiments, varying size, orientation, or brightness, asking the following questions: What set characteristics are perceived? Mean, plus variance, or range? Are these perceived automatically, when they are not part of the performed task? How precise are mean and range perception? How far from the mean need an object be to be considered not the mean? How far from the edge of the range need an object be to be considered out-of-the-range? Finally, we ask if these characteristics are similar for the different tested features.

Methods

Participants

Thirty-nine participants were tested; all reported normal or corrected-to-normal vision. Fifteen participants, students at the Hebrew University of Jerusalem (students), were tested in our laboratory (age range = 20–27 years, mean = 23.7 years; nine men, six women). In addition, we tested 24 participants from Amazon Mechanical Turk (MTurk), a crowdsourcing platform enabling coordination of online participants of up-loaded human information tasks, using Adobe Flash. There was no significant difference between the results of the two groups. All participants provided informed consent and were compensated for participation. All participants were naïve as to the purpose of the experiment.

Apparatus

The laboratory procedure took place in a dimly lit room, with participants seated 50 cm from a 24-in. Dell LCD monitor (Dell, Xiamen, China). All stimuli were shown against a gray background (RGB 0.5, 0.5, 0.5). Stimuli were generated using Psychtoolbox version 3 for MATLAB 2015a (MathWorks, Natick, MA). It is more difficult to know exact physical conditions for MTurk participants.

Stimuli and procedure

We used a RSVP method to present low-level feature stimuli in the center of the display. The in-house student experiment was separated into three sessions with a short break between them, and each session consisted of three blocks for the different low-level feature stimulus types, as shown in Figure 2:

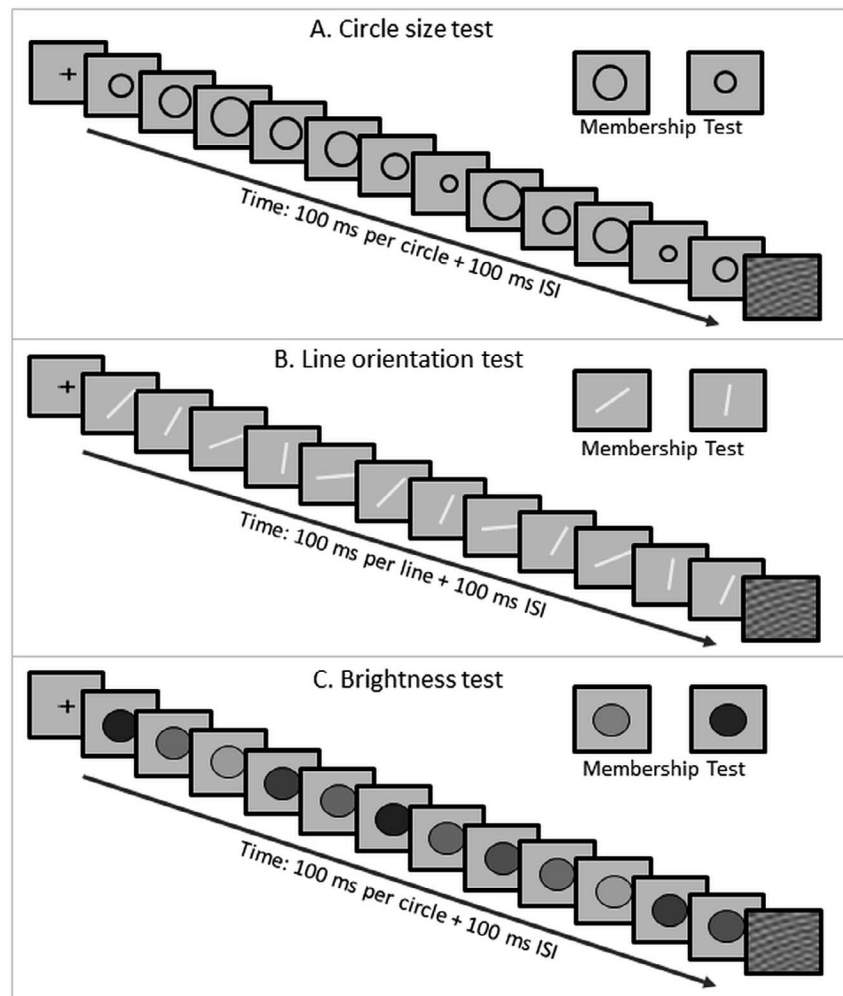


Figure 2. Experimental paradigm. An RSVP sequence of stimuli was presented followed by two test stimuli. Participants chose which test stimulus was a member of the set. (A) Circle size (B) Line orientation (C) Circle brightness. Test element subtypes are detailed in Table 1.

1. Size: 12 black (RGB 0, 0, 0) hollow circles (rings of 0.135 cm width) of six different sizes (two circles for each size) were randomly picked for each RSVP trial, presented in random order, followed by a masking stimulus. The only difference between circles was their varying diameter. The circles database contained 30 different size circles, equally spaced from 0.324 cm minimum to 9.72 cm maximum diameter. Each increment in size (0.335 cm) is termed a “unit.” The range of each trial set was restricted to a maximum of 15 units (4.86 cm), leaving margins of another fifteen sizes out of the range. On each RSVP trial, the range was randomly determined, giving different ranges from trial to trial, between eight and 15 units.
2. Orientation: 12 light gray lines (RGB 0.7, 0.7, 0.7; 8.64 cm × 0.135 cm) of six different orientations (two lines for each orientation in each 12-element

sequence) were randomly chosen for each RSVP trial, presented in random order. The line database contained 30 differently oriented lines, between 0° and 174° with increments of 6°, termed a “unit.” The maximum range of the set was 15 units (84°), leaving margins of another 15 orientations out of the range. On each RSVP trial, the range was randomly determined between 8 and 15 units.

3. Brightness: Twelve 13.5 cm diameter disks with black rings (RGB 0, 0, 0) of 0.054 cm width and filled with six brightness levels (two disks for each gray level in each 12-element sequence) were randomly chosen for each RSVP trial, presented in random order. Disks differed only in brightness and the consequent contrast made with the background and surrounding ring. The brightness database contained 30 different levels of brightness from light gray (RGB 0.79, 0.79, 0.79) to dark gray (RGB 0.21, 0.21, 0.21) with increments

Subtype number	Member test element (correct)	Nonmember test element (incorrect)
1	Amean	A
2	A	A
3	A	Amean
4	A	B
5	Amean	B

Table 1. Trial subtypes.

of 0.02. Thus, the brightness level inside the disk was either darker or brighter than the background and appeared as such despite the black ring. Each increment in brightness is termed a “unit.” The maximal set range was 15 units, from darkest to brightest, leaving another 15 brightness levels out of the range. On each RSVP trial, the range was randomly chosen, between eight and 15 units.

Each session consisted of 60 trials, 20 of each stimulus block, four of each of the five subtypes (see below). Trials were preceded by a fixation cross in the middle of the screen for 2 s, and then 12 elements were displayed sequentially 5/s, i.e., with 200 ms stimulus onset asynchrony (SOAs) consisting of 100 ms stimulus and 100 ms interstimulus interval (ISI). After the RSVP sequence, a 2-AFC membership test was performed with two test elements presented on either side of the screen—one, a member of the trial set and the other a nonmember. Participants were instructed to press the right/left arrow button on the keyboard corresponding to the position of the member element. Each of the three blocks was repeated three times, giving 180 trials total per participant. Participants began with a training session containing 15 trials, five for each stimulus type.

All trial stimulus and exposure time parameters were the same for the MTurk experiment. However, each session (total 3) tested a single stimulus feature, size, orientation, or brightness. Each session had 40 trials, eight for each subtype, making 120 trials total per participant.

Since participants are better at perceiving and memorizing early and late elements, (primacy and recency effects; Hubert-Wallander & Boynton, 2015), the choice of test member elements excluded the first and last two elements of the RSVP sequence. Background color and stimulus position were constant so that only one feature changed during the trial (Figure 2).

Trial subtypes

The experiment was designed with five different trial subtypes (Table 1), pseudorandomly mixed in each

session. The difference between subtypes is only in the membership test stimuli. Each test element could be one of the following options: A—an element in the range of that trial’s RSVP set; Amean—an element equal to that trial’s set mean (size, orientation, or contrast); B—an element out of the set range (necessarily a nonmember test element). There were five trial subtypes in total, numbered as presented in Table 1, without participants being aware of this division.

Statistical tests and data analysis

To verify that performance accuracy and reaction times (RT) depend on test stimulus subtype, we conducted repeated measure two-way analyses of variance (ANOVA) with within-subject factors of mean and range. Additionally, a one-way repeated-measure ANOVA tested differences between subtypes 1, 2, and 3, testing the mean effect isolated from “outside range” subtypes 4 and 5. To investigate significant mean and range effects for specific subtypes, we performed one-tailed *t* tests. We use one-tailed *t* tests because in each case we test significance of a specific direction of influence, where the opposite direction is counterintuitive. Nevertheless, all effects found to be significant, would be significant also under two-tailed *t* tests, as well. To test effects at higher resolution, we plot accuracy and response time (RT) results as a function of distance from set mean (or edge).

Results

We evaluated participant performance by measuring accuracy rates (choosing the element that was a member of the set) and RT for each trial subtype separately. The first step in the analysis was averaging the results across participants to yield a group mean result for each subtype. Figure 3 depicts percentage of correct responses for Students (Figure 3A), MTurk participants (Figure 3B), and all participants combined (Figure 3C), for each of the five trial subtypes. Results for each of the three low-level stimulus blocks, testing size, orientation, and brightness are also shown separately (Figure 3D). The proportion of correct responses for each subtype was relatively close for the three stimulus blocks.

Entering the accuracy rates (of all 39 participants) into a 2 (member element: mean vs. non-mean) \times 2 (nonmember element: in range vs. outside range) repeated measure ANOVA revealed significant main effects of mean, $F(1, 38) = 11.65, p < 0.01$, and range, $F(1, 38) = 345.33, p < 0.001$, and an interaction effect, $F(1, 38) = 14.56, p < 0.001$, as well. The interaction is

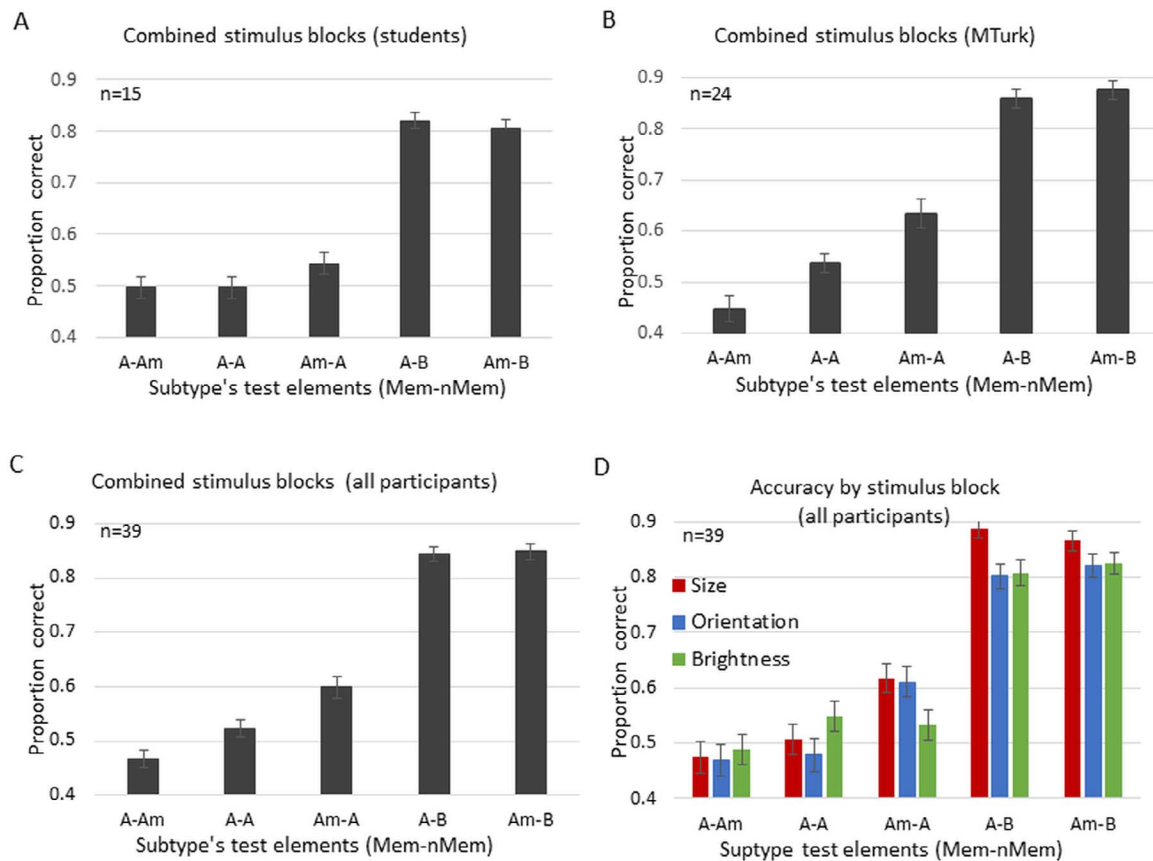


Figure 3. Accuracy by trial subtype. (A–C) Accuracy averaged over stimulus block vs. trial subtype (subtype 1:Am-A; 2:A-A; 3:A-Am; 4:A-B; 5:Am-B; Am = Amean, Mem = member, nMem = nonmember). (A) Students. (B) MTurk. (C) All participants. (D) Accuracy for each stimulus block, (tested dimension), separately (all participants). Error bars, here and in subsequent figures, represent standard errors across subjects for each trial subtype.

due to the mean effect being present only when the non-member test element was in the set range, as shown by post-hoc t tests (see the following materials, following section). We suggest that the absence of a mean effect in “out of range” subtypes was due to its being much smaller than the range effect and a ceiling effect for out-of-range nonmembers. We conclude that variance in accuracy results was due to different subtype difficulties, rather than participants’ arbitrary performance.

Mean effect

The hypothesis behind searching for the mean effect in this membership test is that in the absence of individual element representations, participants would tend to perceive the mean element as a set member. Without being able to remember and recognize member elements, participants may choose the test element closer to the mean, irrespective of whether it was a member of this trial’s set. Indeed, a significant difference was found ($p < 0.001$, for all participants) between trial subtypes with mean feature as the correct

test element with respect to subtypes with “non-mean” feature as the correct test element, as shown in Figure 4 for the student (Figure 4A), MTurk (Figure 4B) and combined results (Figure 4C). This was found for each stimulus block and tested varying dimension, size, orientation, and brightness (Figure 4D).

To zoom into the mean effect, we tested only subtypes presenting two test elements within the set range (subtypes 1, 2, and 3), excluding trial subtypes containing irrelevant (out-of-range) factors and leaving mean element position (member, nonmember or none) as the only variable. We hypothesize that results will reveal a decrease in performance for subtype 3 (A-Amean) due to the nonmember being the mean, and an increase in performance for subtype 1 (Amean-A) due to the member being the mean, compared to subtype 2 (A-A) which lacks the mean at all. Thus, subtype 2 serves as control for this analysis. Indeed, a one-way repeated measure ANOVA (for all 39 participants) on accuracy rates between these subtypes revealed a significant effect of mean presence, $F(1, 38) = 18.93$, $p < 0.001$. Results showed better performance for subtype 1 (Am-A) versus subtypes 2 or 3 (A-A or A-

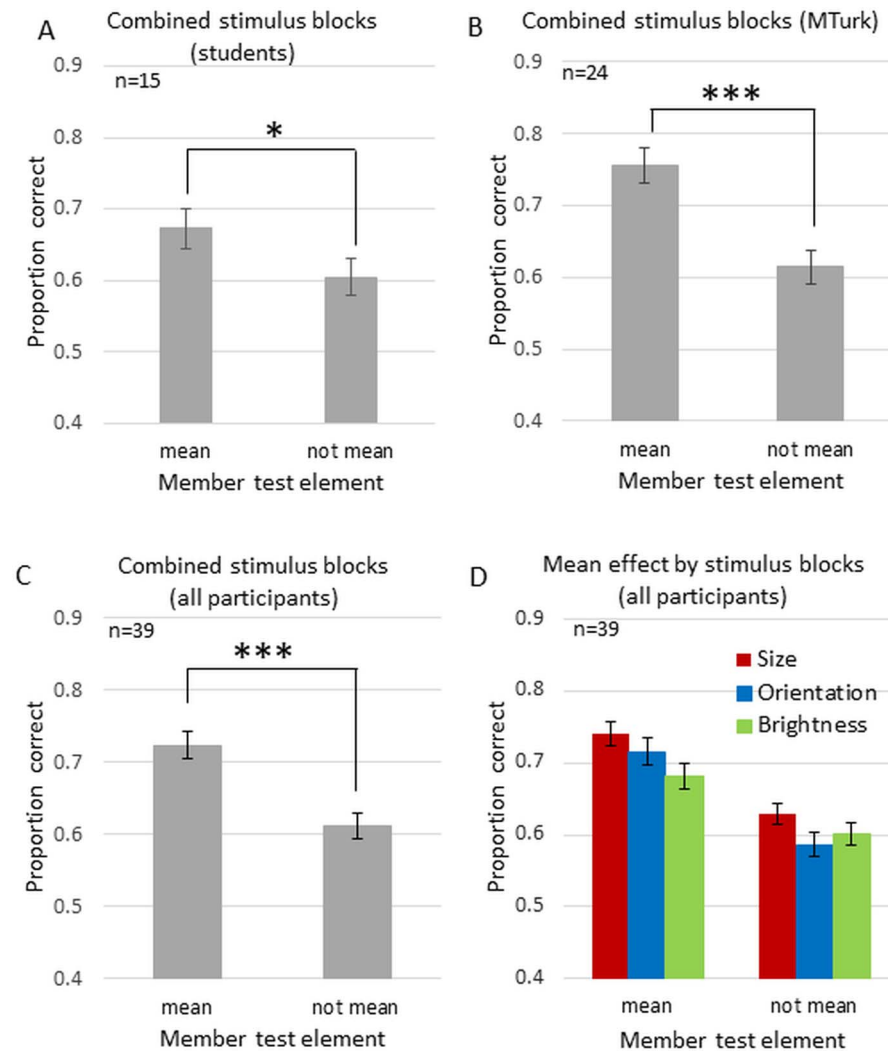


Figure 4. Accuracy results showing mean effect. Accuracy rates for subtypes with member test element equal to the set mean (1, 5) and for subtypes with member test element not the mean (2, 3, 4) combining stimulus blocks. (A) Students. (B) MTurk. (C) All participants. (D) Accuracy rates for each stimulus block separately (all participants). * $p < 0.05$, *** $p < 0.001$.

Am) in each stimulus block (t test: $p < 0.002$) except brightness (Figure 3C). While subtypes 2 and 3 were surprisingly quite similar for the student group (3A and 6A), for MTurk participants (3B and 6B) the effect was significant between each of these subtypes (t test: $p < 0.005$), as predicted.

For a more specific analysis regarding the influence of proximity to the mean, we calculated accuracy as a function of the test element (member/nonmember) relative distance from the mean of the RSVP set (and not merely whether they are exactly the mean feature). Figure 5 illustrates this effect as analyzed by the data of subtype 2 (A-A), showing that accuracy decreases as the member test element is farther from the mean feature (Figure 5A), and accuracy increases as the non-member test element is more distant from the mean (Figure 5B). Gray circles in Figures 5A and 5B are taken from subtypes 3 (A-Amean) and 1 (Amean-A),

respectively, in which one of the test elements is exactly the mean feature. A gradual trend in accuracy is seen within subtype 2 (A-A), with the most accurate trials being those with the member closest to the mean and the non-member the most distant, as demonstrated by the three-dimensional plot of Figure 5C.

We measured the isolated mean effect (subtypes 1, 2, 3, disregarding subtypes 4 and 5 with out-of-range nonmember) as a function of the difference between the distances of the test elements from the mean. To what extent do participants choose the element that is closer to the mean? Figure 6A and 6B illustrates participant accuracy performance for each participant group. Results show a mild decrease in performance if the non-member is the set mean (subtype 3, A-Amean compared to subtype 2, A-A; Students: n.s.; MTurk: $p < 0.01$); a significant increase if the member is the set mean (subtype 1, Amean-A, compared to subtype 2, A-

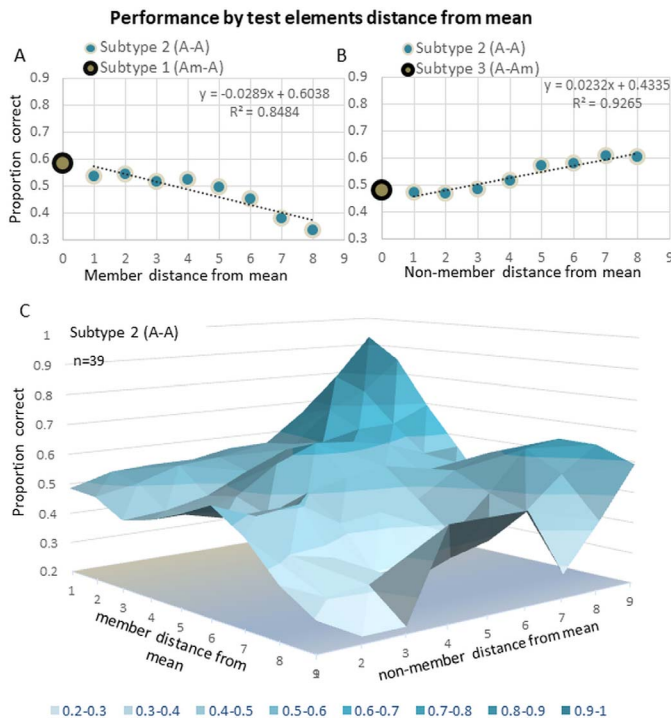


Figure 5. Graded mean effect with distance from mean (all participants). Subtype 2 (A-A) accuracy as function of member test element distance from set mean (A) and as function of nonmember test element distance from set mean (B). Gray circles represent accuracy for (A) subtype 1 (Am-A) or (B) subtype 3 (A-Am). (C) Performance accuracy for subtype 2 (A-A) as a function of distance of both test elements from the mean.

A; Students: $p < 0.05$; MTurk: $p < 0.01$); and a highly significant effect comparing presence of the set mean as the member or the non-member (subtype 3, A-Am), compared to subtype 1, A-mean-A; both groups: $p < 0.001$). Note that performance for the control subtype 2 (A-A), where neither test element is the set mean, is near chance level (0.50 ± 0.03 and 0.53 ± 0.01 for Students and MTurk, respectively). Performance is significantly beyond chance when the member is the mean (0.60 ± 0.04 and 0.63 ± 0.02) and below chance when the nonmember test element is the mean (0.48 ± 0.03 and 0.44 ± 0.02). This suggests that the dominant effect for response is not element memory but rather perception and memory of the set mean. This is especially striking since set mean is determined on the fly for each trial separately.

In Figure 6C and 6D we plot performance as a function of the difference between the distances of the test elements from the mean, when the member was closer to the mean (right side of the graphs) or further from it (left side). This is plotted for each subtype separately (blue, red, and gray curves for subtypes 1, 2, and 3, respectively) and averaged together (green), for Students (Figure 6C) and MTurk participants (Figure 6D). The black dashed line is the linear regression of

the averaged three subtypes, showing improved performance as the member test element is closer to the mean, and poorer performance as the non-member element is closer to the mean. These results suggest that observers tend to respond relying not only on their perception of the set mean, but also on the proximity of the test elements to it.

(C-D) Accuracy as function of difference between test elements' distances from the mean for Students (C) and MTurk (D). On the left, the nonmember test element is closer to the mean; on the right, the member test element is closer to the mean. Green line is average performance of all three subtypes; dashed line is its trendline.

Range effect

We also investigated perception of the second set characteristic—representation of its boundaries, termed the “Range Effect.” Improvement in performance on the membership test when the nonmember deviates from the RSVP set boundaries would lead to the conclusion that not only the set mean but also the set range is represented. This is demonstrated first in Figure 3, and in more detail in Figures 7 and Figure 8. As long as both test elements were in the set range (subtypes 1, 2, and 3), average performance was poor (0.52 ± 0.01 proportion correct). In contrast, subtypes (4 and 5) with out-of-range nonmember elements showed highly significant superior performance (0.83 ± 0.01 correct; $p < 0.001$; Figures 7A and 7B).

We divided the accuracy results for out-of-range subtypes (4 and 5) into trials where the nonmember element was higher versus lower than the set range, to test if the range effect is present for both sides. In both cases, performance was high (0.85 ± 0.01 and 0.82 ± 0.01 correct for nonmember below the minimum and above the maximum, respectively). Comparing these results to the in-range trial subtypes (1, 2, and 3), we found a significant performance improvement for each side ($p < 0.001$), as seen in Figure 7B. Response time measurements also revealed a significant difference ($p < 0.001$) between the in-range and out-of-range subtypes (1.20 ± 0.02 s nonmember test element in-range; 1.06 ± 0.02 s out-of-range), suggesting that out-of-range judgments are made more rapidly (Figure 7C). Taken together, these results suggest that observers' representation of the range of a set included the boundaries on both sides of a set's variable feature.

In Figure 8 we plot results for the nonmember out-of-range subtypes (4 and 5) as a function of the distance of the non-member element from the set range closer edge (Figure 8A: Students; 8B: MTurk). There is a strong upward slope reflecting a gradual increase in

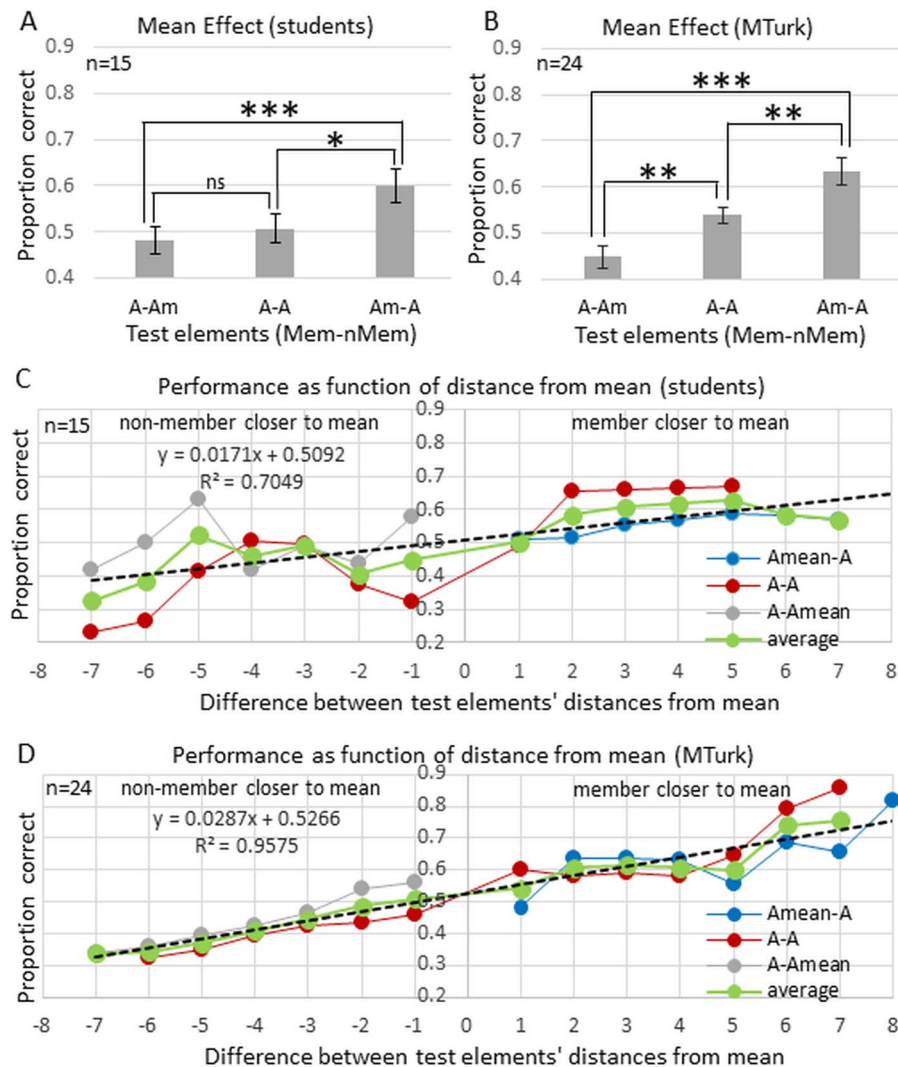


Figure 6. Membership test performance as function of mean presence or distance from test elements on in-range subtypes (1, 2, and 3). (A–B) Accuracy rates for each subtype and *t* tests between them for students (A: excluding trials with <4 units distance) and MTurk (B). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

performance as the nonmember test element is more distant from the set range, indicating a gradual range effect (Students slope = 0.023, $R^2 = 0.893$; MTurk: slope = 0.019, $R^2 = 0.75$).

Combining data from the two participant groups and comparing subtypes 2 (A–A) and 4 (A–B), we plot in Figure 9A a continuous presentation of the range effect, as a function of nonmember test element distance from the mean, whether in- or out-of-range. This analysis reveals a jump in performance when crossing out of the range, as indicated by the arrow in Figure 9. Accuracy improves as the distance from the mean increases, with a jump between regression lines of in- (gray dots) and out- (blue dots) of-range. An ANOVA with main factor non-member in or out of range (A–A vs. A–B) for nonmember distance from mean 5–8, revealed that this jump is significant, $F(1, 6) = 21.325$, $p < 0.01$, and similarly when the member

equals the mean (Amean–A vs. Amean–B; $p < 0.05$). This jump can only be due to a perceptual representation of the range of the set, not only of its mean, and not even of mean plus variance. Taken together, these results lead to the conclusion that observers had a sharp representation of the range of a set, which included the boundaries on both sides of the set's variable feature.

The range effect can also be seen independent of the mean effect, as demonstrated in Figure 9B. Here, we plot performance accuracy for subtypes A–A and A–B, that is comparing performance when the nonmember element was within or outside the sequence range, as in Figure 9A. However, here we plot performance for each stimulus value, independent of the position of the range and of the mean value. There is a striking difference between the two

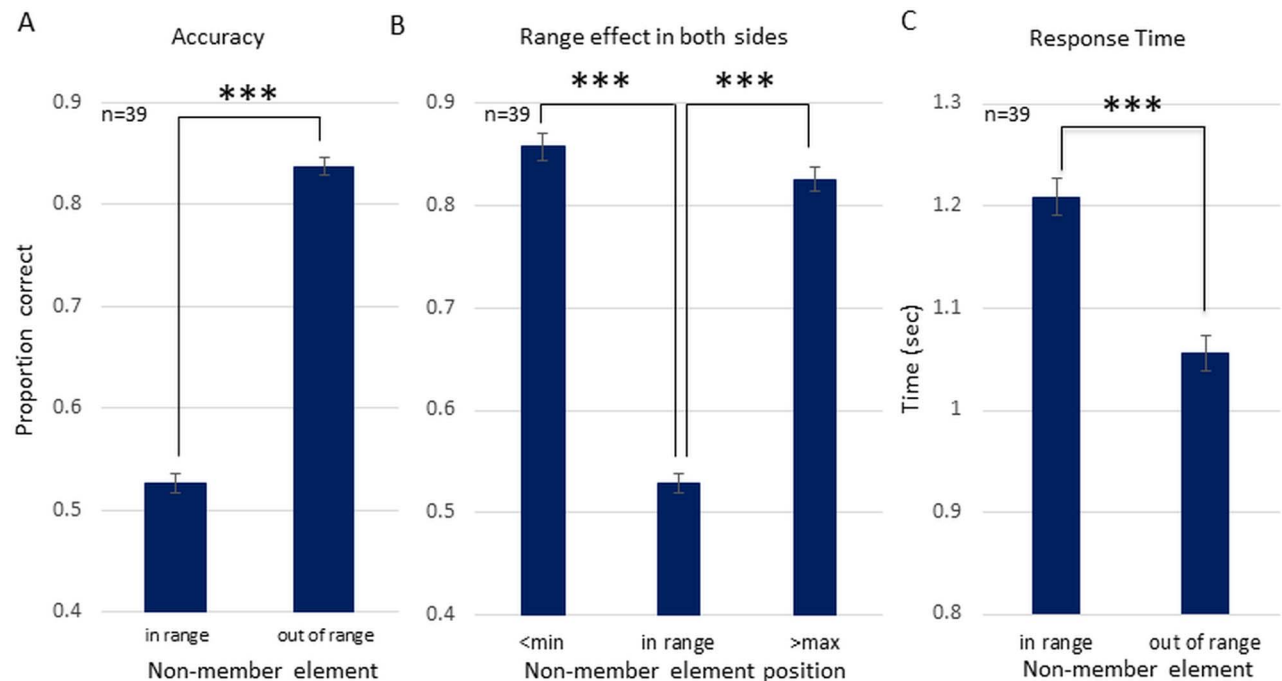


Figure 7. Range effect: accuracy and response time (all participants). (A) Accuracy for averaged subtypes from both nonmember test element forms. (B) Accuracy rates for trials where the nonmember element is either lower, higher, or in the range of the set. Results from subtypes 4 and 5 for “>max” and “<min” trials, and subtypes 1, 2, and 3 for “in range” trials. *t* tests showed highly significant differences for each side of the range. (C) Response time for in- and out-of-range subtypes. *** $p < 0.001$.

curves, indicating that observers were well aware of the sequence range.

Following the experimental session, we queried student observers as to their strategy in performing the task. All responded that they tried as hard as possible to recall the absolute stimulus values in the sequence, to determine which test item was a member of the set. The only exception was that some indicated that for extreme sets, (e.g., all items very large or very small), they were able to choose as a member the test item that matched the set.

Summary and discussion

In summary, regarding basic features of size, orientation, and brightness, two main phenomena were found by a 2-AFC experiment testing memory of an element out of a sequence of 12: (1) Mean effect—the mean representation-biased participant decision toward choosing the element closer to the mean as the member of the RSVP sequence, whether it was or was not actually the member. (2) Range effect—the range representation assisted participants in rejecting elements outside the RSVP sequence range, therefore correctly choosing the member element quite easily and achieving high performance with these trials. Importantly, participants were never informed about the

involvement of mean, range, or any other statistical properties, and they only followed the instructions of a simple visual memory task. Thus, any use of information as to the mean and range of the set must come from automatic, implicit perception of these statistical properties.

We confirmed the mean effect for three perceptual features, size, orientation, and brightness and found that participants perceive the mean online, i.e., for each trial’s set of stimuli independently, and implicitly, and without set mean information being relevant to the performed task. Presence of the RSVP set mean as member or nonmember in the 2-AFC test, impacted strongly on membership judgments (Figures 3, 4), and proximity to the mean was effective gradually (Figures 5, 6). Experimental conditions were intentionally designed, with multiple stimuli presented with short SOAs, to make it difficult for participants to memorize all the elements, leading to their relying on their mean representation. Participants depended only on implicit perception of the mean (and range) and chose the test element closer to the mean, to the extent that performance is below chance level when the nonmember is closer to the mean. Previous results on ensemble processing found that observers can process up to four items every 210 ms (Gorea, Belkoura, & Solomon, 2014), suggesting that in our study observers had sufficient time to process individual elements. Nevertheless, performance is quite close to chance

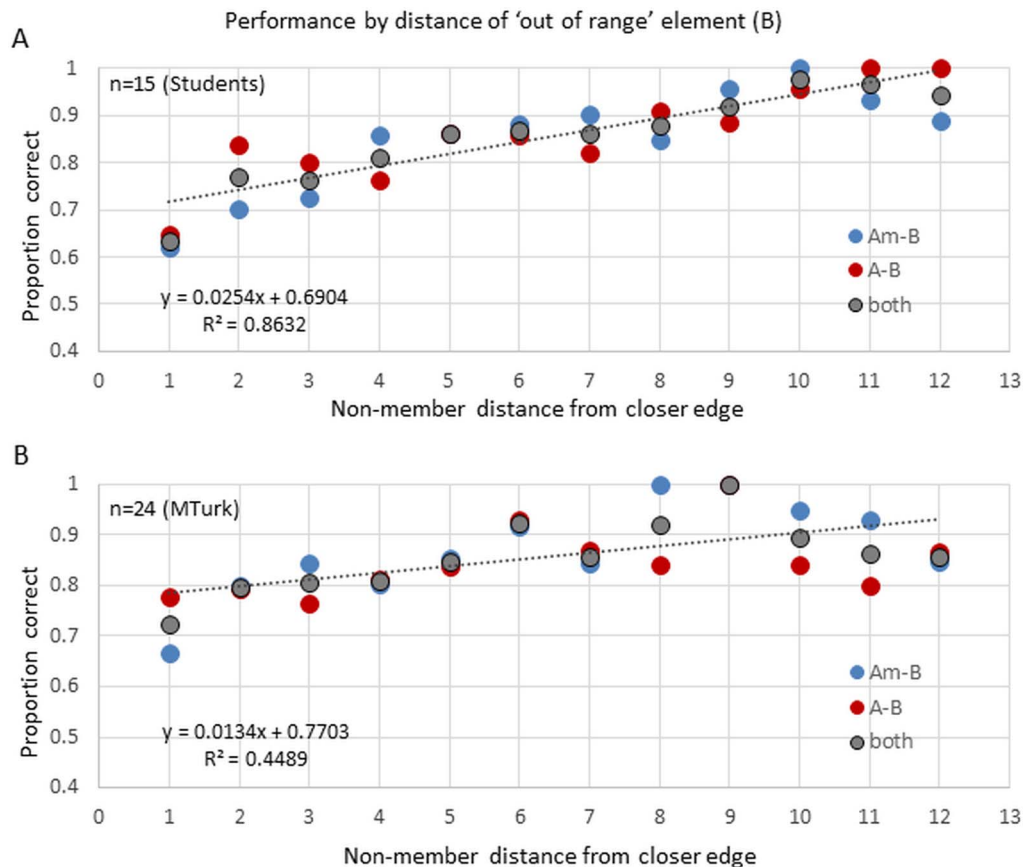


Figure 8. Range effect: increased accuracy as the non-member element more distant from set's edge in (A) Students. (B) MTurk. The red dots represent trial subtype 4 and the blue dots represent trial subtype 5. The gray dots connected by the black line represent the average of both subtypes. Dashed-line represents the trendline of the average of both subtypes.

level when both test elements are within the set range. This indicates that even when observers had time to process individual elements, the surprising result is that, nevertheless, they do not succeed to remember all presented elements, but rather depend on their implicit perception of the sequence mean. We do not know what would have been the outcome had the participants been able to view each stimulus at leisure (Standing, 1973). Comparing results as a function of test element distance from the mean, we found a gradual shift in accuracy with respect to the member or nonmember element distance from the mean (Figures 5, 6, 9). Thus, the ensemble mean is perceived, though approximately. This characterization correlates with previous results of visual summary statistics (Haberman & Whitney, 2009) and the intuitive system of numerical averaging, as well (Brezis et al., 2015).

The second, less discussed, characteristic perceived is the set range—a sequence with a variable feature from one boundary to another. Results of Ariely (2001) already suggest that observers exhibit some knowledge of the range of sets, excluding from their membership judgments test elements beyond the set's

edges. As shown in Figures 7–9, observers succeeded in telling if one of the test elements was out of the size, orientation, or brightness range, therefore choosing the other element as a member. In these cases, where one test element is outside the set range, performance jumps from close to chance to ~80% correct. There is a well-known perceptual phenomenon called “the central tendency of judgment”, where observers tend to judge elements as closer to the mean than they really are, in a range-dependent manner (Hollingworth, 1910). Although this may be an independent phenomenon, it, too, shows that observers have an implicit knowledge of the mean and range of sequences of stimuli.

These capabilities of our perceptual systems are widely used in our daily encounters with objects, without awareness of the mechanisms nor of their effect on behavior. We see a tree full of leaves and roughly perceive average and range of their size, color, density, etc. We represent highways by summaries of overall speed and direction of motion, only secondarily by properties of each vehicle. The ability to include summarized properties from multiple items serves perception efficiency, saving time and energy required

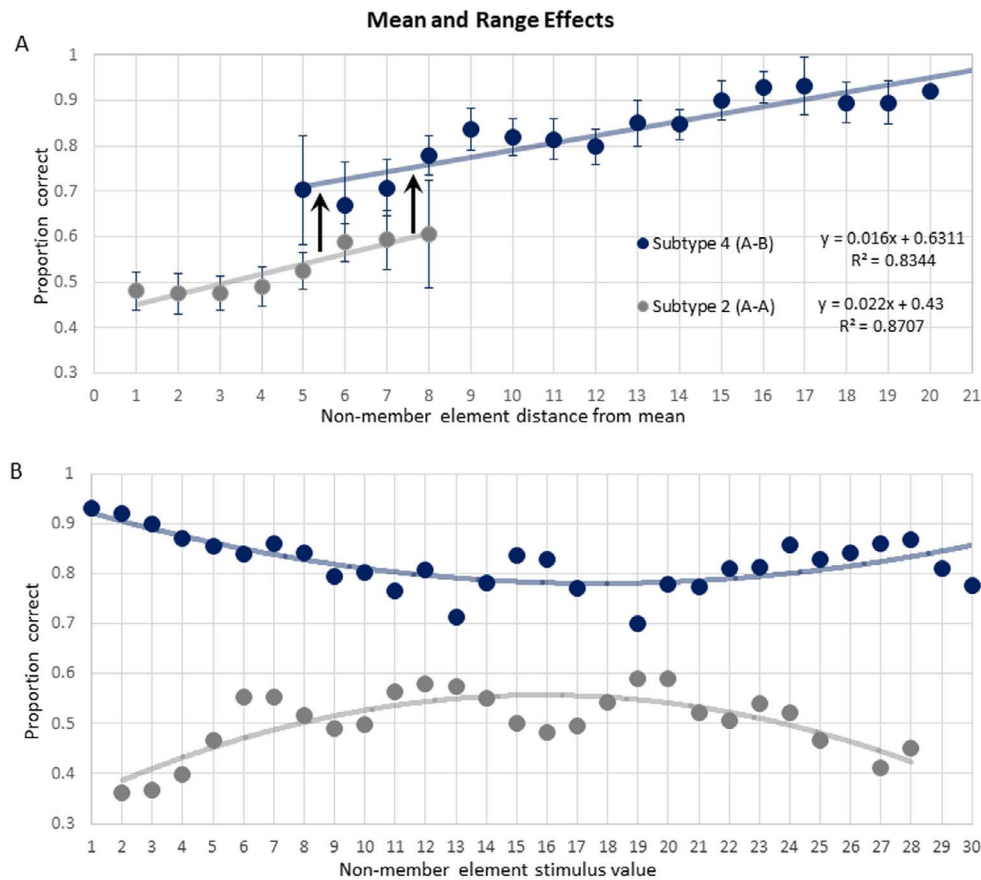


Figure 9. (A) Nonmember element distance from set mean effect on performance in and out of range (Students). Gray dots represent the accuracy rates for each distance of the nonmember test element from the mean, averaged over all stimulus blocks in subtype 2 (A-A). Blue dots represent the accuracy rates as function of distance from mean when the non-member is out-of-range, averaged over all stimulus blocks in subtype 4 (A-B). Black arrow indicates the “jump” of the trendline as observed after nonmember crosses out of range. (B) Range effect as function of absolute stimulus value, irrespective of distance from mean or position of range.

to encode information of small elements, separately. Range representation, too, serves daily perception and behavior by isolating elements whose features deviate from the range of sets of scene elements, so that they pop out in an automatic detection process (Rosenholtz, 1999; Treisman & Gelade, 1980).

We leave to future research discovery of the cerebral sites and cortical mechanisms underlying set summary statistic perception, including mean and range perception.

Keywords: summary statistics, mean, range, visual dimensions

Acknowledgments

We thank Yuri Maximov, the lab’s talented programmer, and lab co-members Safa’ Abassi and Miriam Carmeli. Thanks to Merav Ahissar, Udi Zohary, Israel Nelken, Robert Shapley, Howard Hock,

and Anne Treisman (of blessed memory), for helpful discussions of earlier drafts of this paper.

This study was supported by a grant from the Israel Science Foundation (ISF).

Commercial relationships: none.

Corresponding author: Shaul Hochstein.

Email: shaulhochstein@gmail.com.

Address: Life Sciences Institute and ELSC Safra Center for Brain Research, Hebrew University, Jerusalem, Israel.

References

- Albrecht, A. R., Scholl, B. J., & Chun, M. M. (2012). Perceptual averaging by eye and ear: Computing summary statistics from multimodal stimuli. *Attention, Perception, & Psychophysics*, 74(5), 810–815.

- Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2014). Obligatory averaging in mean size perception. *Vision Research*, *101*, 34–40.
- 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, USA*, *106*(18), 7345–7350.
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162.
- Attarha, M., & Moore, C. M. (2015). The capacity limitations of orientation summary statistics. *Attention, Perception, & Psychophysics*, *77*(4), 1116–1131.
- Bauer, B. (2009). Does Stevens's power law for brightness extend to perceptual brightness averaging? *The Psychological Record*, *59*(2), 171.
- Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports*, *5*.
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404.
- Chong, S. C., & Treisman, A. (2005). Attentional spread in the statistical processing of visual displays. *Attention, Perception, & Psychophysics*, *67*(1), 1–13.
- Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, *20*(5), 324–335.
- Corbett, J. E., & Oriet, C. (2011). The whole is indeed more than the sum of its parts: Perceptual averaging in the absence of individual item representation. *Acta Psychologica*, *138*(2), 289–301.
- Cowan, N. (2001). Metatheory of storage capacity limits. *Behavioral and Brain Sciences*, *24*(1), 154–176.
- Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192.
- Fabre-Thorpe, M. (2011). The characteristics and limits of rapid visual categorization. *Frontiers in Psychology*, *2*.
- Gorea, A., Belkoura, S., & Solomon, J. A. (2014). Summary statistics for size over space and time. *Journal of Vision*, *14*(9):22, 1–14, <https://doi.org/10.1167/14.9.22>. [PubMed] [Article]
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), R751–R753.
- Haberman, J., & Whitney, D. (2009). Seeing the mean: ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–784.
- Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception, & Psychophysics*, *72*(7), 1825–1838.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804.
- Hollingworth, H. L. (1910). The central tendency of judgment. *Journal of Philosophy, Psychology and Scientific Methods*, *7*(17): 461–469
- Huang, Y., & Rao, R. P. (2011). Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(5), 580–593.
- 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, 1–12, <https://doi.org/10.1167/15.4.5>. [PubMed] [Article]
- Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. A. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition*, *48*, 149–160.
- Luck, S. J., & Vogel, E. K. (1997, November 20). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.
- Maule, J., & Franklin, A. (2015). Effects of ensemble complexity and perceptual similarity on rapid averaging of hue. *Journal of Vision*, *15*(4):6, 1–18, <https://doi.org/10.1167/15.4.6>. [PubMed] [Article]
- McDermott, J. H., Schemitsch, M., & Simoncelli, E. P. (2013). Summary statistics in auditory perception. *Nature Neuroscience*, *16*(4), 493–498.
- Morgan, M., Chubb, C., & Solomon, J. A. (2008). A 'dipper' function for texture discrimination based on orientation variance. *Journal of Vision*, *8*(11):9, 1–8, <https://doi.org/10.1167/8.11.9>. [PubMed] [Article]
- Neumann, M. F., Schweinberger, S. R., & Burton, A. M. (2013). Viewers extract mean and individual identity from sets of famous faces. *Cognition*, *128*(1), 56–63.
- Piazza, E. A., Sweeny, T. D., Wessel, D., Silver, M. A., & Whitney, D. (2013). Humans use summary statistics to perceive auditory sequences. *Psychological Science*, *24*(8), 1389–1397.
- Robitaille, N., & Harris, I. M. (2011). When more is

- less: Extraction of summary statistics benefits from larger sets. *Journal of Vision*, *11*(12):18, 1–8, <https://doi.org/10.1167/11.12.18>. [PubMed] [Article]
- Rosenholtz, R. (1999). A simple saliency model predicts a number of motion popout phenomena. *Vision Research*, *39*(19), 3157–3163.
- Solomon, J. A. (2010). Visual discrimination of orientation statistics in crowded and uncrowded arrays. *Journal of Vision*, *10*(14):19, 1–16, <https://doi.org/10.1167/10.14.19>. [PubMed] [Article]
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, *25*(2), 207–222.
- Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(2), 329–337.
- Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.
- Utochkin, I. S. (2015). Ensemble summary statistics as a basis for rapid visual categorization. *Journal of Vision*, *15*(4):8, 1–14, <https://doi.org/10.1167/15.4.8>. [PubMed] [Article]
- Ward, E. J., Bear, A., & Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals?: The role of statistical perception in determining whether awareness overflows access. *Cognition*, *152*, 78–86.
- Yamanashi Leib, A. Y., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*, 13186.