

Supplemental Materials for “Not All Summary Statistics Are Made Equal: Evidence from Extracting Summaries Across Time”

Bjorn Hubert-Wallander & Geoffrey M. Boynton

Here are presented five additional behavioral experiments that further measured the temporal weighting profile of mean position and mean size judgments. The primary purpose of the experiments was to test the reproducibility of the findings from Experiment 1 in the main paper. Namely, finding primacy for judgments of mean dot location and finding recency for judgments of mean dot size. The secondary purpose of the experiments was to measure the effect of various experimental manipulations on these temporal weighting profiles. The method in each experiment was very similar to that used in Experiment 1 in the main paper, and so only deviations from that method are noted here.

Supplemental Experiment 1: Averaging Position with Feedback

To understand how estimates of mean dot location are affected by the presence of corrective feedback, we presented participants with a series of small white dots and asked them to report the mean position of the group. Some participants were presented with the actual correct answer alongside their answer after their report, while others were not. We then quantified the influence of the temporal position of each dot on the participants' location estimates across many trials in each group of participants.

Method

Participants

Two groups of sixteen undergraduate students at the University of Washington were recruited and compensated in the same manner as in Experiment 1 in the main paper. Participants were assigned to their group at the time of the experimental session by referring to a randomly pre-shuffled condition list.

Stimuli

Each trial consisted of a series of ten small white dots that varied in their position. Stimuli were the same as in Experiment 1 in the main paper with the exception that here all dots were the same size, their radii equal to 0.3° of visual angle. Dot locations were sampled as previously described, except that here the standard deviation of the sampling distribution on each trial was 3.0° .

Procedure

Experimental procedure was the same as in Experiment 1 in the main paper with the following exceptions. While the central cross that signaled the start of each trial was the same as it was previously, no verbal precue or postcue was used here since only one stimulus property (mean location) was being reported here. Additionally, a feedback period of 1500 ms was added after the response period of each trial. In one group of participants (the no-feedback group), the feedback period consisted of only a green dot indicating the participant's response. In the other group (the feedback group), the feedback period contained a green dot indicating the participant's response and also a red dot indicating the actual, correct answer. After the feedback period, the next trial began immediately.

Each participant completed 400 trials in ten blocks of 40 trials each. In addition to the normal task instruction, the experimenter also explained the dots contained in the feedback period to participants in both groups, and encouraged both groups to try to answer as accurately as possible on each trial.

Results

Weights describing the relative influence of each temporal position in the sequence of dots (one through ten) on each participant's responses were obtained by fitting the weighted average model described in the main text. Average weights by group are shown in Figure S1.

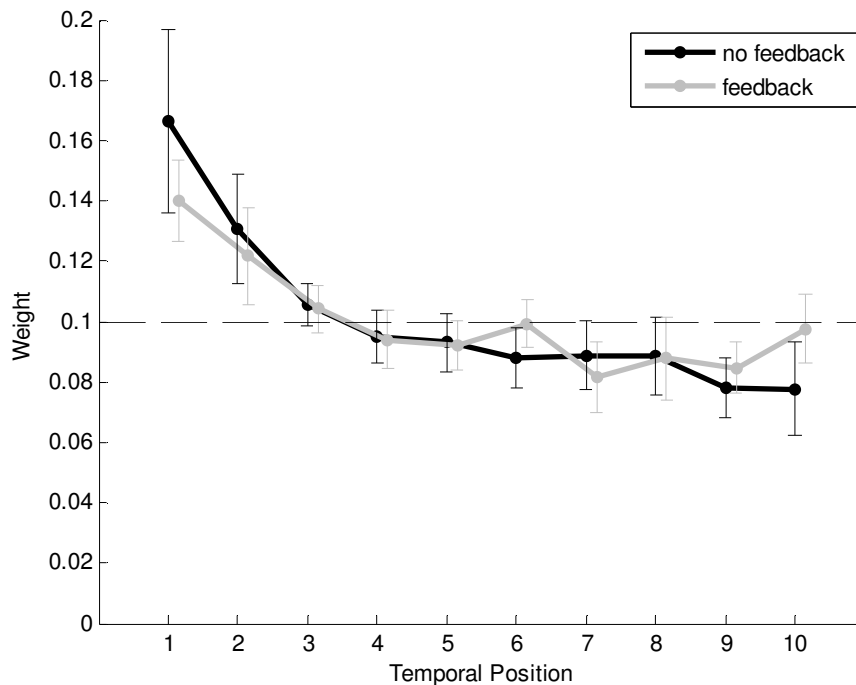


Figure S1. Results from Supplemental Experiment 1. Mean weights as a function of temporal position per group for judgments of mean position. X-axis indicates the temporal order of the dots shown, where 1 refers to the first dot presented on each trial and 10 refers to the last dot presented on each trial. Y-axis indicates the relative influence of each dot on participants' responses in each group. Weights were obtained by fitting a weighted average model to the data (see Results text for details). Dashed line indicates the weights expected if all dots contributed equally to responses and no other source of noise or bias were present. Weights from the feedback condition have been displaced rightward for readability. Error bars indicate 95% confidence intervals across participants.

As in Experiment 1 in the main text, primacy was found for judgments of mean dot location. Earlier dots appeared to contribute more to participant responses than later dots did, a finding supported by a significant main effect of temporal position in a two-way repeated measures ANOVA, $F(9,270) = 20.27, p < 0.001, \eta_p^2 = .40$. Though there appeared to be no overall difference in the magnitude of weights in the feedback and no-feedback group, the main effect of feedback condition was statistically significant, $F(1,30) = 6.57, p = 0.02, \eta_p^2 = .18$. However, there was no significant interaction between temporal position and feedback condition, $F(9,270) = 1.54, p = 0.14, \eta_p^2 = .05$, suggesting that the presence of feedback after each trial did not affect the shape of the temporal weighting profile compared to no feedback.

Supplemental Experiment 2: Averaging Position at Different Presentation Rates

Here we manipulated the rate at which dots were presented within each participant, adding in trials with half the presentation rate as that used in Experiment 1 in the main text.

Method

A new group of sixteen participants was recruited as before. Apparatus, stimuli, and procedure were as described in Supplemental Experiment 1 with the following exceptions.

Stimuli

The only difference in stimuli between Supplemental Experiment 1 and this one was that here dots were shown at two different presentation rates across trials. In “fast” trials, each dot was present for 150 ms and was followed by a 50 ms blank inter-dot interval. In “slow” trials, each dot was present for 300 ms and was followed by a 150 ms blank inter-dot interval. It should be noted that “fast” trials used the same presentation rate (5 Hz) as in Experiment 1, but that “slow” trials used half that rate (2.5 Hz).

Procedure

There was no feedback to participants in this experiment, with the feedback period being replaced by a 1500 ms blank inter-trial interval. Participants completed 400 trials each, in ten blocks of 40. Each block contained either all fast trials or all slow trials, with fast and slow blocks alternating over the course of the experiment. First block type encountered was random and counterbalanced across participants. Participants were explicitly told about the different presentation rates, and encountered the same number of both types during the practice phase, which consisted of about ten trials.

Results

Weights were obtained for each participant for each condition in the manner described previously. Mean weights across participants in each presentation rate condition are shown in Figure S2.

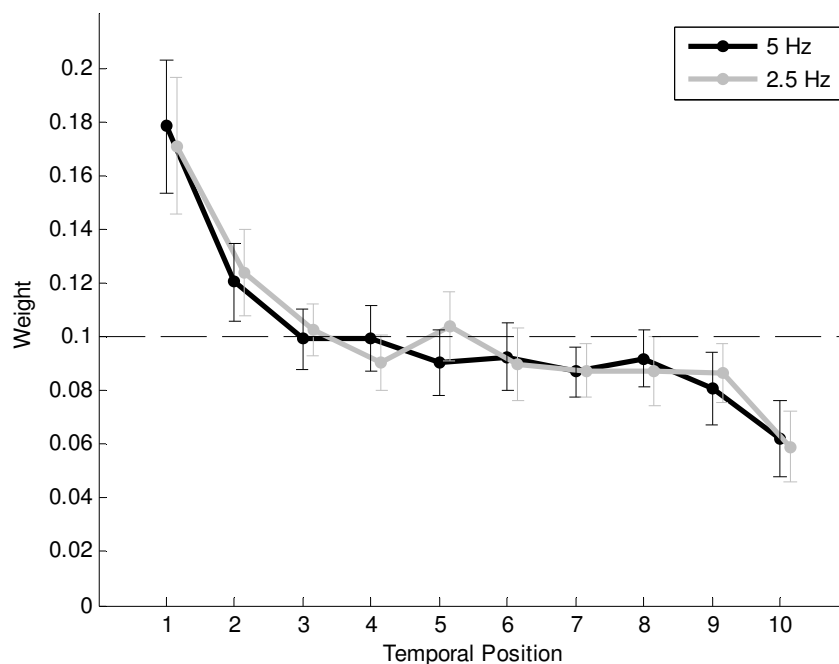


Figure S2. Results from Supplemental Experiment 2. All figure details are same as in Figure S1.

Again an unmistakable pattern of primacy was found for estimates of mean dot location, evidenced by a significant main effect of temporal position, $F(9,135) = 25.56$, $p < 0.001$, $\eta_p^2 = .63$. Interestingly, the pattern of weights was indistinguishable across the two presentation rates, with no significant main effect of

presentation rate, $F(1,15) = 0.78$, $p = 0.3$, $\eta_p^2 = .05$, or interaction between presentation rate and temporal position, $F(9,135) = 0.53$, $p = 0.85$, $\eta_p^2 = .03$. Giving the participants more time per dot to update their belief about the mean position of the dots did not appear to affect the temporal weighting profile of their responses at all. Primacy was still found.

Supplemental Experiment 3: Averaging Position With More or Less Variable Dots

Instead of manipulating presentation rate within participants, in this experiment we manipulated the average spatial spread of the dots to see if this would the temporal weighting profile the participants applied to the sequence of dots in judging mean position.

Method

A new group of seventeen participants was recruited as before. Apparatus, stimuli, and procedure were as described in Supplemental Experiment 2 with the following exceptions.

Stimuli

Two trial types were used in this experiment. All stimulus parameters were the same between trial types except for the standard deviation of the sampling distribution that dot positions were sampled from.

“Narrow” trials used a standard deviation of 1.5° of visual angle and “wide” trials used a standard deviation of 4.8° . All trials used a presentation rate of 5 Hz, just as in Supplemental Experiment 1 and in Experiment 1 in the main text.

Procedure

Participants completed 400 blocks of 40 trials each, with the two trial types randomly intermixed throughout all blocks. Participants saw an equal number of narrow and wide trials in the practice phase, which consisted of about ten trials.

Results

Weights were obtained for each participant for each condition in the manner described previously. Mean weights across participants in each location variance condition are shown in Figure S3.

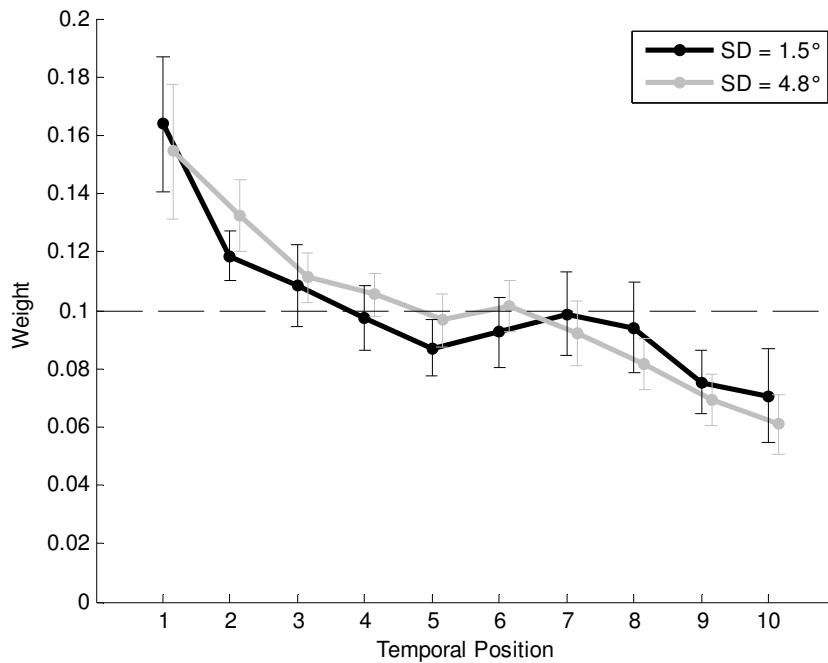


Figure S3. Results from Supplemental Experiment 3. All figure details are same as in Figure S1.

Once again, an overall pattern of primacy was found, with earlier dots contributing more to participants' responses than later dots. This observation was supported by a significant main effect of temporal position in a two-way ANOVA, $F(9,144) = 21.02, p < 0.001, \eta_p^2 = .57$. However, changing the mean spatial spread of the dots shown on each trial did not appear to affect the weights participants applied to the dots, with no significant main effect of the condition variable, $F(1,16) = 0.31, p = 0.58, \eta_p^2 = .02$, and no significant interaction between variables, $F(9,144) = 1.67, p = 0.10, \eta_p^2 = .09$.

Supplemental Experiment 4: Averaging Size

The purpose of this experiment was to test for the presence of recency in computing summaries for mean size across time when dot position did not vary and only size judgments were being made.

Method

Participants

A new group of fifteen undergraduate students at the University of Washington was recruited and compensated as described in Experiment 1 of the main text.

Stimuli

On each trial in this experiment, participants viewed a series of eight white dots that varied only in their size. Each dot was shown for 200 ms and was followed by a 133 ms blank inter-dot interval, resulting in a dot presentation rate of 3 Hz. On a given trial, dot radii were chosen by sampling eight times from a Gaussian probability distribution, the center of which was sampled on each trial from a uniform distribution ranging from 0.6° to 2.3° of visual angle. The standard deviation of the Gaussian distribution was always set to 0.3 times the center of the same distribution in order to minimize the possibility of sampling dot radii below zero. If a dot's radius was sampled to be below zero, it was resampled until it was above zero.

The stimulus area and the response area were separated in distance, with the stimulus dots appearing 5° to the left of the center of the screen and the response dot (which only appeared after the stimulus dots) appearing 5° to the right of the center of the screen. A pair of grey vertical hash marks marking the horizontal position of these locations was visible throughout the experiment so that participants knew where to expect stimulus and response dots. Each hash mark in a pair was located 5° vertically away from the horizontal meridian of the screen such that even large dots rarely came within 1° of the mark.

Procedure

Each trial began with the presentation of a white cross approximately 0.8° in width and height at the dot presentation location. After 250 ms, the cross became red to signal the imminent arrival of the dot sequence. After a random period between one and two seconds, the cross disappeared and was immediately followed by the dot sequence. The dot sequence was followed by a 600 ms blank period, which was followed by the response period. Here the participants were expected to report the "average size" of the dots shown on that trial by adjusting a sample response dot. The response dot's initial radius was chosen randomly on each trial from 0.1° to 3.2° , which also served as the limits of possible responses. The response period ended when the participant submitted his or her response, and was followed by a 500 ms inter-trial interval.

Each participant completed 350 trials and was given opportunity to take breaks after every 40 trials, though participants were also advised that they could take a break at any time by waiting to submit their response for a given trial. Each participant received full instructions from an experimenter and then completed approximately ten practice trials in view of the experimenter before beginning the experimental trials. A full experimental session lasted about an hour.

Results

Weights were obtained for each participant for each condition in the manner described previously. Mean weights across participants are shown in Figure S4.

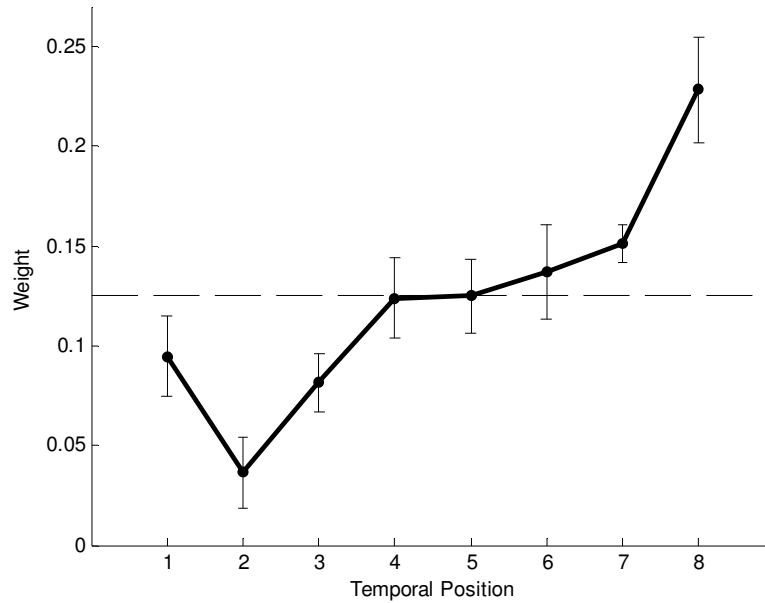


Figure S4. Results from Supplemental Experiment 4. Mean weights as a function of temporal position for judgments of mean size. All other figure details are same as in Figure S1.

As in Experiment 1 in the main text, we found recency for computing mean size across time, where the later dots appeared to contribute more to responses than earlier dots. This was supported by a significant effect of temporal position in a one-way repeated measures ANOVA, $F(7,98) = 29.08$, $p < 0.001$, $\eta_p^2 = .68$. Though the overall pattern of recency likely contributed significantly to this outcome, the statistical significance was likely also partially driven by the effect seen in the first three dots. Interestingly, the first dot appeared to suffer less downweighting than the second, giving the temporal weighting profile a checkmark-like appearance that was not observed in Experiment 1 in the main text.

Supplemental Experiment 5: Averaging Size at Different Presentation Rates

In this final experiment, we manipulated the rate of dot presentation and measured the effect on temporal weighting profiles for judgments of mean size.

Method

A new group of twenty participants were recruited and compensated as described previously. Apparatus, stimuli, and procedures were the same as in Supplemental Experiment 4 with the following exceptions.

Stimuli

The stimuli used here were exactly the same as used in Supplemental Experiment 4, excepting that there were two types of trials in this experiment. In “fast” trials, each dot was present for 133 ms and was followed by a blank inter-dot interval of 67 ms, resulting in a presentation rate of 5 Hz. In “slow” trials, each dot was present for 333 ms and was followed by a blank inter-dot interval of 167 ms, resulting in a presentation rate of 2 Hz.

Procedure

Participants each completed 320 trials in eight blocks of 40. Each block contained either all fast trials or all slow trials, with fast and slow blocks alternating over the course of the experiment. First block type encountered in a given session was random and counterbalanced across participants. Participants were explicitly told about the different presentation rates, and encountered the same number of both types during the practice phase, which consisted of about ten trials.

Results

Weights were obtained for each participant for each condition in the manner described previously. Mean weights across participants in each presentation rate condition are shown in Figure S5.

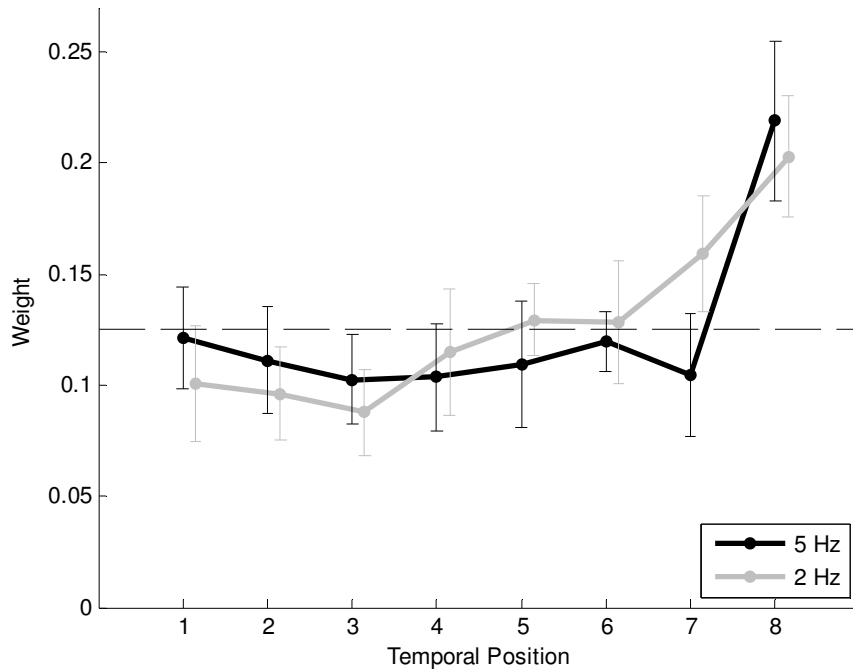


Figure S5. Results from Supplemental Experiment 5. Mean weights as a function of temporal position for judgments of mean size for each trial type. All other figure details are same as in Figure S1.

As in Supplemental Experiment 4 and Experiment 1 in the main text, we found a general pattern of recency for estimates of mean dot size across time. This was supported by a significant main effect of temporal order in a two-way repeated measures ANOVA, $F(7,133) = 13.77, p < 0.001, \eta_p^2 = .42$. Additionally, a significant main effect of presentation rate was also found, $F(1,19) = 9.80, p < 0.01, \eta_p^2 = .34$. Interestingly, inspection of the temporal weighting profiles suggests that weights increased gradually over the course of a trial in the 2 Hz presentation rate condition, while weights remained relatively constant until the last dot in the 5 Hz presentation rate condition, where the weight increased abruptly. However, the interaction between temporal order and presentation rate only approached significance, $F(7,133) = 1.96, p = 0.07, \eta_p^2 = .09$, providing no evidence that varying presentation rate affects the shape of the temporal weighting profile.

General Discussion

The primary purpose of the experiments presented here was test the reproducibility of the basic findings of Experiment 1 in the main text. Results from all five supplemental experiments did exactly that, consistently

finding primacy for estimating mean dot position and recency for estimating mean dot size across a variety of conditions.

The secondary purpose of these experiments was to measure the effect of various stimulus manipulations on the temporal weighting profiles in both position and size domains. In the position domain, we manipulated presentation rate, the variance of the underlying sampling distribution, and incorporated corrective feedback, but none of these conditions lead to a change in the temporal weighting profile participants applied to the stimuli, at least to the extent discernable by our analysis method and statistical power. In the size domain, manipulating presentation rate produced two weighting profiles that appeared to differ in shape, but the effect was not statistically significant. In the main text, we raise the point that summary representations computed across time may differ across feature domains, so the possibility that weighting profiles in the size domain might be more malleable than those in the position domain is intriguing. However, our supplemental results do not allow strong conclusions to be made on this point, since we were not able to significantly or consistently perturb the pattern of weights participants applied to the sequences of dots in either position or size domains. Further study will be needed to explore how manipulable temporal weighting profiles are in various feature domains.