The Impact of Spatial and Temporal Regularities on Cognitive Processes: A Preliminary Study within the Scope of the EU Project

Introduction

Humans excel at statistical learning (SL), the ability to detect regularities that help guide attention in noisy environments (Nobre & van Ede, 2018; Võ et al., 2019). In visual search, this has been demonstrated by contextual cueing: repeated spatial configurations speed up target detection compared to novel ones (Chun & Jiang, 1998; Sisk et al., 2019). This effect reflects configural learning of distractor–target relations, where long-term memory templates guide attention toward the target location. However, temporal regularities can also facilitate performance. For example, fixed distractor sequences or temporal intervals can predict target onset and improve detection (Olson & Chun, 2001; Cravo et al., 2017).

Most prior work, however, has studied spatial and temporal SL separately, leaving open how participants exploit regularities when information unfolds simultaneously across space and time. To address this, Experiment 1 tested whether search is facilitated when full spatiotemporal sequences—including distractor identities, spatial positions, and temporal order—are repeated across trials, compared to randomly varying sequences.

Experiment 1: Repetitions of complete sequences

Methods

Participants

A group of 25 participants (mean age = 26, SD = 6.78, females = 16) naïve to the purpose of the study took part in Experiment 1. Because no prior work has tested SL with multiple spatial, temporal, and visual regularities, no direct effect size estimate was available for formal power analysis. We therefore recruited N=25 participants, which is at the upper end of comparable SL studies (that typically tested between 10 and 20 participants, e.g., Olson & Chun, 2001; Correa et al., 2005; Rohenkohl et al., 2014; Denison et al., 2017). A sensitivity analysis (paired-samples t, two-tailed, α=.05) conducted in R using the *pwr* package (Champely, 2020) shows that this sample size provides 80% power to detect medium effects of dz = 0.58. For comparison, a metaanalysis of spatial SL studies reported an average RT-based contextual-cueing effect size of d = 1.0 (Vadillo, Konstantinidis, & Shanks, 2016), suggesting that our sample size is conservative relative to established contextual facilitation. Participants were tested individually in a quiet and dimly lit room. All participants provided written informed consent before participation and received either course credit points or monetary compensation (10€). The study protocol was approved by the LMU ethics committee and was conducted following the Declaration of Helsinki.

Apparatus

All experimental stimuli were presented on a 23-inch ASUS LCD monitor (VG236H model; 1920 × 1080 pixels; 60 Hz refresh rate). The monitor was controlled by a Windows 10 PC running MATLAB R2022b and Psychtoolbox 3.0.18 (Brainard, 1997; Kleiner et al., 2007). Viewing distance was precisely fixed at 80 cm using an adjustable chin rest to ensure consistent retinal stimulation. Participant responses were collected via a standard QWERTY keyboard. The monitor's display area measured 51 cm in width and 29 cm in height, resulting in a pixel density of 0.0266 cm per pixel. All stimuli were presented against a uniform black background ([0, 0, 0] RGB).

Stimuli and Display

Throughout the experiment, five square placeholders (100 × 100 pixels, or 2.66 × 2.66 cm, subtending approximately 1.91° × 1.91° of visual angle) were shown in fixed positions, equidistantly spaced on an imaginary circle (radius: 300 pixels = 7.98 cm = 5.71°) centered on the screen. Placeholders were centered at polar angles of 0°, 45°, 90°, 135°, and 180°, corresponding to right, upper-right, top, upper-left, and left relative to fixation. The center-to-center distance between adjacent placeholders was 353 pixels (9.39 cm, ≈6.71° of visual angle). All placeholders were mid-gray ([127, 127, 127] RGB), and a white fixation cross ([255, 255, 255] RGB) was displayed at the screen center throughout each trial.

Procedure

On each trial, a sequence of five symbols was presented: four uppercase letters (randomly drawn from the English alphabet) and one digit (4 or 5, randomly assigned per trial). Symbols were rendered in black ([0, 0, 0] RGB), with a font size of 50 pixels (height ≈1.33 cm ≈ 0.95° visual angle), and appeared individually, centered within their designated placeholder. At any given moment, only one symbol was visible, while the remaining placeholders remained empty.

The experiment consisted of 100 blocks, each containing 10 trials (five “repeated” and five “random” trials), yielding 1,000 trials in total. Short, enforced breaks were provided after every five blocks to minimize fatigue. Each trial began with a 500-ms fixation interval, during which all five placeholders and the central fixation cross were displayed. Next, five symbols (four letters and one digit) were presented sequentially, one at a time, each for 350 ms, with each symbol appearing in its assigned placeholder; the fixation cross remained onscreen throughout. Only one symbol was displayed at any time, with the rest of the placeholders remaining empty. Upon presentation of the digit, participants had up to 2,000 ms to respond via keypress, indicating whether the digit was a 4 or 5. Response mapping (i.e., which arrow key corresponded to each digit) was counterbalanced across participants (4 = up or down; 5 = down or up). Participants were instructed to respond as quickly and accurately as possible, ignoring the letter symbols. Feedback was provided for 1,000 ms following each response window. Correct responses triggered a blank interval, whereas incorrect or missed responses elicited the word “Wrong!” in red ([255, 0, 0] RGB).

Figure 1a shows the schematic arrangement of the five placeholders. In Figure 1b, we illustrate one possible repeated sequence with target temporal order (TTO) = 4: The distractor letter “M” appears at the upper-right → distractor “P” at the top → distractor “F” at the left → the target digit “4” at the right → and the distractor “S” at the upper-left. In this repeated context, the distractor identities, their order, and their spatial locations remain invariant across repetitions, while the target identity (“4” vs. “5”) is randomized across trials (to prevent stimulus-response learning, e.g., “This context means “4”). For a different repeated context with, e.g., TTO = 1, the distractor set, their locations, and their order were different (but held constant across repetitions) —for example: a target “4” appearing at the top → a distractor letter “K” at the left → distractor “A” at the upper-right → distractor “D” at the right → and distractor “B” at the upper-left. Likewise, for TTO = 5, a different repeated context could have this sequence: distractor “H” at the upper-left → distractor “R” at the right → distractor “C” at the top → distractor “T” at the upper-right → target “4” at the left. Thus, across the five repeated contexts, each TTO and each target location occurred exactly once.

Experimental Conditions

Each repeated context had a unique combination of distractor identities, their spatial locations, and their temporal order, associated with a fixed target temporal position and a fixed target location. Thus, the exact distractor sequence defining one repeated context never appeared in any other repeated context. In repeated trials, this configuration of four letters and one target digit was fixed and presented 50 times throughout the experiment. For each participant, five unique repeated contexts were generated such that, across these five contexts, the target digit appeared once at each of the five possible temporal positions (positions 1–5) and spatial locations (one per placeholder). This assignment ensured that no two repeated contexts shared the same temporal or spatial target position; thus, across all contexts, each of the five temporal and spatial target positions was uniquely represented. The same constraints were applied to random contexts, ensuring that target temporal and spatial positions were equally represented, but with distractor identities, orders, and locations varying unpredictably from trial to trial. On each block, five repeated and five random trials were presented in randomized order.

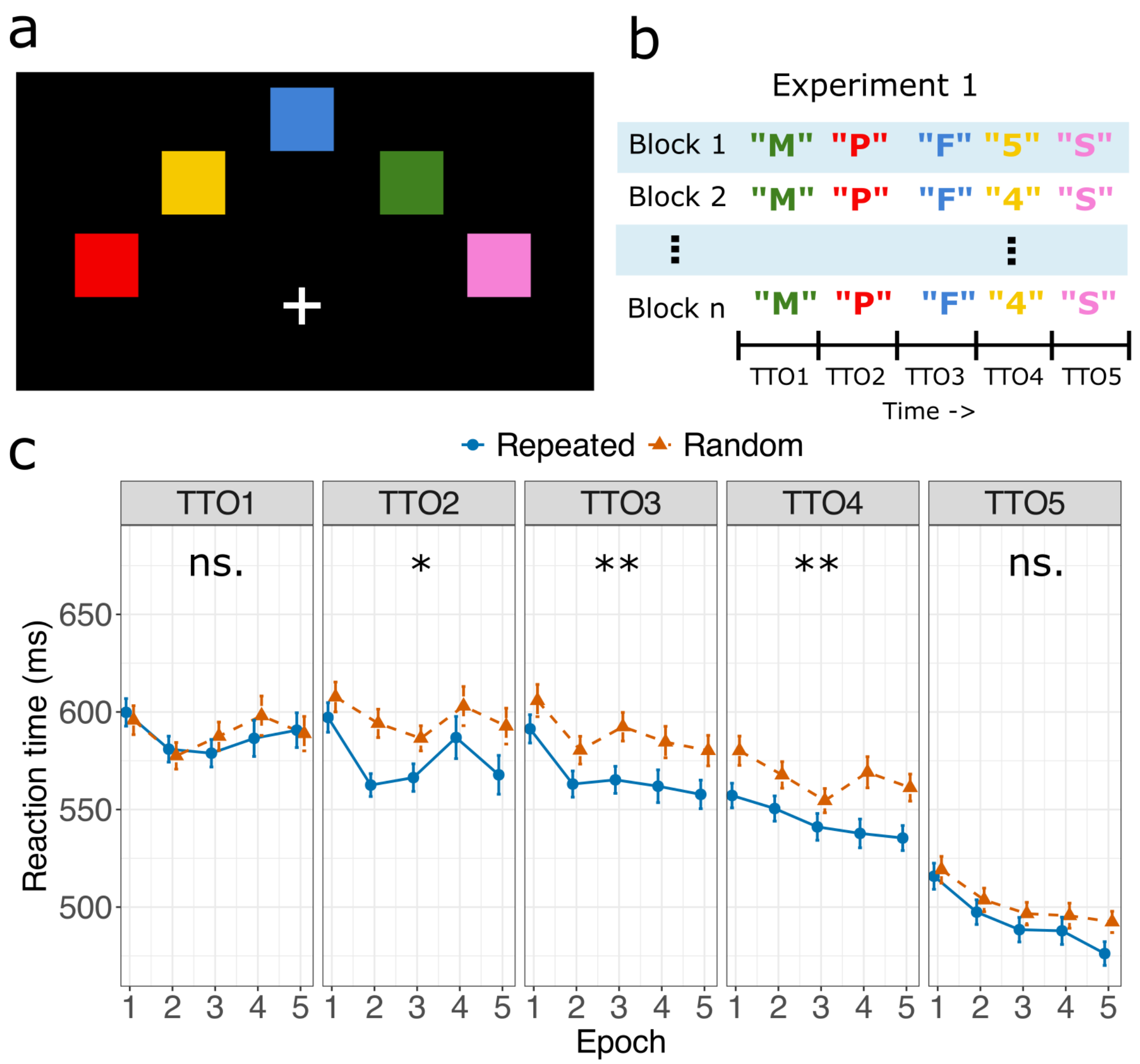

Figure 1. (a) Setup standard to all five experiments: participants sat in a dimly lit and quiet room, viewing five placeholders arranged in a semicircle on a monitor. (b) Example of a repeated trial from Experiment 1 (target temporal order, TTO #4). Letters indicate distractor identities; colors mark their respective placeholder locations. In the actual task, all placeholders were gray; colors are used here solely for illustration. (c) Mean reaction times (RTs) for each TTO, averaged across 25 participants, shown by epoch (20-block segments). Error bars represent the standard error of the mean (SE). Statistical comparisons for TTOs #1 to #5 were conducted by pooling across epochs and contrasting repeated with non-repeated trials. Evidence for contextual cueing effects is indicated by Bayes factors (BF_{10}): n.s. = $BF_{10} < 3$; * = BF_{10} 3–10 (moderate evidence); ** = $BF_{10} > 10$ (strong evidence).

During the main task, participants were given a 2,000-ms response window, which began at the onset of the target digit, regardless of its temporal position within the sequence. The target digit was displayed for 350 ms and then replaced by the following symbol in the sequence (or a blank screen, if it was the fifth and final position). When the target digit appeared at temporal positions 1 through 4, the remaining symbols in the sequence were presented as usual, and participants could respond at any point within the 2,000-ms interval following target onset. If the target digit appeared in the final (fifth) position, the screen remained blank for the remainder of the response window (i.e., for 2,000 ms minus 350 ms), during which participants could make their response.

Statistical analysis

Data were analyzed with linear mixed-effects models (LMMs; lme4, Bates, Mächler, Bolker, & Walker, 2015) including fixed effects of Context (Repeated vs. Random), Target Temporal Order (TTO 1–5), and Epoch (1–5), with random intercepts for participants. Models were fit to participant-cell means (Subject × Context × TTO × Epoch). Our *primary analyses* consisted of Bayesian paired-samples t-tests at each TTO (averaging over Epoch), as Bayes factors provide a direct quantification of evidence for contextual cueing (rather than a dichotomous significance decision) and remain interpretable without multiplicity correction. These analyses were implemented with the BayesFactor package (Morey & Rouder, 2023) using a JZS Cauchy prior (scale = √2/2). Reported BF_{10} values quantify evidence for the alternative over the null and are interpreted following van Doorn et al. (2021). To provide a direct link to the classical contextual cueing literature and to facilitate comparison, we also report complementary frequentist statistics in Table 2. Specifically, we obtained estimated marginal means (emmeans; Lenth, 2023) for planned Repeated vs. Random contrasts at each TTO, reported Cohen's dz effect sizes, and adjusted p-values for multiplicity across the five TTO-wise contrasts using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995; FDR q = .05). Effect sizes from the LMM ANOVA are reported as partial η^2 (computed with the 'effectsize' package; Ben-Shachar, Lüdecke, & Makowski, 2020). All analyses were performed in R 4.3.2. Finally, because contextual cueing is typically expressed in response times rather than error rates, and since response accuracy was uniformly moderate across all experiments (repeated = 6%, SD = 8.60, random = 6.32%, SD = 8.75), we report detailed error analyses only in Supplement Table 2 (results of all linear mixed-effects models across all experiments) and Supplement Table 3 (results of all planned comparisons). In brief, error rates showed position- and practice-related effects but no reliable context effects and thus did not qualify the RT findings. This reporting strategy was applied consistently across Experiments 1–5.

Results

The results of Experiment 1 are summarized in Figure 1, and mixed-effects model summaries for all experiments are provided in the Supplementary Table 1. Further, Table 2 below summarizes all exact values, including effect sizes, p-values (raw and BH-corrected), and Bayes factors, from the planned comparisons. As expected, manual reaction times were faster for repeated compared to non-repeated context, an effect that could be seen at TTO-2 (19.73 ms; d=-.63, BF_{10}=9.67), TTO-3 (21.15 ms, d=-.66, BF_{10}=13.98), and TTO-4 (21.59 ms, d=-.63, BF_{10}=10.14), but not with TTO-1 (3.04 ms; d=-.09, BF_{10}=.23) and TTO-5 (10.34 ms, d=-.42, BF_{10}=1.34). From these results, it appears that participants could utilize repeated sequences and form reliable expectations about the search for the target, at least at TTO-2 to TTO-4, possibly because these sequences had repetitions of all types of contextual regularities—the identities of the distractor letters, their spatial positions, and their temporal order. – The lack of contextual cueing at TTO-1 is not surprising, given that the target was the first stimulus in the sequence. Thus, additional search-guiding context information was revealed only later in the trial and could aid performance only during later TTOs. By contrast, the absence of a contextual benefit at TTO-5 is surprising, given that in this condition the target was the last item in the sequence and thus could maximally benefit from contextual repetitions. However, our sequences had only a limited number of (five) stimuli, where a target was always present in each trial. Therefore, there was an increasing probability that participants would expect the target in later time windows: TTO-1 (20%); TTO-2 (25%); TTO-3 (33%); TTO-4 (50%); TTO-5 (100%), which may have curtailed the contextual facilitation effect. Under this account, the absence of a contextual facilitation effect at TTO-5 may reflect a floor effect, as the mean RT at this position was already optimal—likely too fast to allow for further measurable improvement (by the repeated contexts; see Figure 1).

Table 1

Planned comparisons for context effects at each target temporal order (TTO) in Experiments 1, 2, 3, 4, and 5.

Exp	TTO	ΔRT (ms)	t	p	pBH	d	BF_{10}	Random M (SD)	Repeated M (SD)
1	1	3.04	0.37	.715	.715	-0.09	0.23	595.66 (154.15)	592.63 (136.46)
1	2	19.73	2.37	.018	.030	-0.63	9.67	599.27 (137.76)	579.54 (153.75)
1	3	21.15	2.54	.011	.028	-0.66	13.98	590.49 (118.81)	569.34 (121.74)
1	4	21.59	2.60	.010	.028	-0.63	10.14	567.16 (107.28)	545.57 (103.82)
1	5	10.34	1.24	.214	.268	-0.42	1.34	499.43 (85.32)	489.09 (97.51)

Note. Exp = Experiment number, ΔRT (ms) = mean difference in reaction time between random and repeated contexts (positive values indicate faster responses for repeated contexts). All p-values are two-tailed. pBH = Benjamini–Hochberg corrected p-value. BF_{10} = Bayes factor in favor of the alternative hypothesis over the null hypothesis.

General Discussion

Experiment 1 shows that contextual cueing extends to sequentially presented displays when spatial, temporal, and identity information are repeated together. Even though the target could appear at any location and time point, participants became faster in repeated contexts, indicating that they could integrate multiple regularities into a single contextual memory trace. This result demonstrates that contextual guidance is not restricted to static displays but can also operate in dynamic event streams, where predictive information unfolds continuously across space and time. In line with earlier work on spatial contextual cueing (Chun & Jiang, 1998) and temporal regularities (Olson & Chun, 2001; Cravo et al., 2017), our findings suggest that spatiotemporal continuity provides a strong scaffold for statistical learning, allowing participants to anticipate where and when to orient attention in complex search environments.