

Spatial and temporal context jointly modulate the sensory response within the ventral visual stream

Tao He^{1,2,3*}, David Richter³, Zhiguo Wang⁴ and Floris P. de Lange³

¹ School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing 100871, China

² IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China

³ Donders Institute for Brain, Cognition and Behaviour, Radboud University, Kapittelweg 29, 6525 EN Nijmegen, the Netherlands

⁴ Institutes of Psychological Sciences, Hangzhou Normal University, Hangzhou, 311121, China

* Correspondence: t.be@wur.nl

12

13 Abbreviated title: Spatial and temporal context predictions

14

15 **Competing Interest:** The authors declare no competing financial interest.

16

17 **Acknowledgements:** This work was supported by the National Natural Science
18 Foundation of China Grant 31371133 to Z.W., The Netherlands Organisation for
19 Scientific Research Vidi Grant 452-13-016 to F.P.d.L., the EC Horizon 2020 Program
20 ERC Starting Grant 678286 “Contextvision” to F.P.d.L, the James S.McDonnell
21 Foundation 220020373 to F.P.d.L and the China Scholarship Council (CSC;
22 201608330264) to T.H.. We thank Wanlu Fu and Zehao Huang for assistance with
23 data collection.

24

25 **Abstract**

26 Both spatial and temporal context play an important role in visual perception
27 and behavior. Humans can extract statistical regularities from both forms of context
28 to help processing the present and to construct expectations about the future.
29 Numerous studies have found reduced neural responses to expected stimuli
30 compared to unexpected stimuli, for both spatial and temporal regularities. However,
31 it is largely unclear whether and how these forms of context interact. In the current
32 fMRI study, thirty-three human volunteers were exposed to object stimuli that could
33 be expected or surprising in terms of their spatial and temporal context. We found a
34 reliable independent contribution of both spatial and temporal context in modulating
35 the neural response. Specifically, neural responses to stimuli in expected compared
36 to unexpected contexts were suppressed throughout the ventral visual stream.
37 Interestingly, the modulation by spatial context was stronger in magnitude and more
38 reliable than modulations by temporal context. These results suggest that while both
39 spatial and temporal context serve as a prior that can modulate sensory processing
40 in a similar fashion, predictions of spatial context may be a more powerful modulator
41 in the visual system.

42

43 **Significance Statement**

44 Both temporal and spatial context can affect visual perception, however it is
45 largely unclear if and how these different forms of context interact in modulating
46 sensory processing. When manipulating both temporal and spatial context
47 expectations, we found that they jointly affected sensory processing, evident as a
48 suppression of neural responses for expected compared to unexpected stimuli.
49 Interestingly, the modulation by spatial context was stronger than that by temporal

50 context. Together, our results suggest that spatial context may be a stronger
51 modulator of neural responses than temporal context within the visual system.
52 Thereby, the present study provides new evidence how different types of predictions
53 jointly modulate perceptual processing.

54 **Introduction**

55 Humans are exquisitely sensitive to visual statistical regularities. Indeed,
56 knowledge of both spatial and temporal context can facilitate visual perception and
57 perceptual decision-making (Bar, 2004). For instance, in the case of spatial context,
58 a foreground object is more easily identified when it appears on congruent
59 backgrounds, compared to when it appears on incongruent backgrounds (Davenport
60 and Potter, 2004). Facilitatory effects of temporal context have also been shown, for
61 instance during exposure to sequentially presented stimuli, with faster and more
62 accurate responses to expected compared to unexpected stimuli (Bertels et al.,
63 2012; Hunt & Aslin, 2001; Richter & de Lange, 2019). At the same time neural
64 responses have been shown to be modulated by temporal context, with a marked
65 suppression of sensory responses to expected compared to unexpected stimuli,
66 reported in humans (Summerfield et al., 2008; den Ouden et al., 2009; Egner et al.,
67 2010; Richter et al., 2018; Richter and de Lange, 2019) and non-human primates
68 (Freedman et al., 2006; Meyer and Olson, 2011; Kaposvari et al., 2018). However,
69 comparatively less is known about the modulation of neural responses by spatial
70 context. Human fMRI studies suggest that a similar network of (sub-)cortical areas is
71 involved in learning spatial contexts as during learning of temporal sequences
72 (Karuza et al., 2017). Thus, while the learning process of temporal and spatial
73 regularities may share neural characteristics, the consequences for sensory
74 processing, following the acquisition of spatial regularities, remain unknown. In
75 particular, do predictions of spatial context result in a similar suppression of neural
76 responses as temporal sequence predictions? Moreover, it is currently unclear if and
77 how spatial and temporal context may interact in shaping sensory processing.

78 In the current study, we set out to concurrently examine the neural and
79 behavioral consequences of spatial and temporal contextual expectations following
80 statistical learning. To this end, participants were exposed to leading image pairs,
81 consisting of two object images presented left and right of fixation, which predicted
82 the identity of trailing object image pairs, thus rendering the trailing images expected
83 based on the temporal context. Moreover, the simultaneously presented images
84 were also predictive of each other, thus generating a predictable spatial context (see
85 **Figure 1c**). Blood oxygenation level-dependent (BOLD) signals were recorded with
86 functional magnetic resonance imaging (fMRI), while participants monitored the
87 images for occasional target images (i.e., flipped object images) that occurred at
88 unpredictable moments.

89 To preview our results, we show that spatial and temporal context both
90 modulate sensory processing in key areas of the ventral visual stream, with
91 pronounced reductions in neural responses to stimuli predicted by spatial and
92 temporal context, compared to stimuli occurring in unexpected contexts.
93 Interestingly, spatial context predictions resulted in a larger suppression than
94 temporal context predictions, suggesting that spatial context may be a more potent
95 modulator of visual processing than temporal context.

96

97 **Materials and Methods**

98

99 **Data and code availability**

100 All data and code used for stimulus presentation and analysis is freely available on
101 the Donders Repository (<https://data.donders.ru.nl/login/reviewer-96936509/hUq0EMV2cQaXlzewhI3XLeHsm3q5xbRMZoSX6-YzhpZc>).

103

104 **Participants**

105 Thirty-three healthy, right-handed participants (13 females, aged 22.36 ± 2.38 years,
106 mean \pm SD) were recruited in exchange for monetary compensation (100
107 Yuan/hour). All participants reported normal or corrected-to-normal vision and were
108 prescreened for MRI compatibility, had no history of epilepsy or cardiac problems.
109 The experiments reported here were approved by the Institutional Review Board of
110 Psychological Sciences at Hangzhou Normal University and were carried out in
111 accordance with the guidelines expressed in the Declaration of Helsinki. Written
112 informed consent was obtained from all participants. Data from two participants were
113 excluded. Of these two exclusions, one participant's behavioral performance of the
114 post-scanning task was at chance level, while the other participant showed
115 excessive head motion (i.e., a number of relatively head motion events exceeding 1
116 mm notably above the group mean).

117

118 **Stimuli**

119 The object images were a selection of stimuli from Brady et al. (2008), and also
120 previously used by Richter and de Lange (2019). A subset of 48 full color object
121 stimuli, comprised of 24 electronic objects and 24 non-electronic objects were shown
122 during the present study. For each participant, 24 objects (12 electronics and 12
123 non-electronics) were pseudo-randomly selected, of which 6 (including 3 electronics)
124 were pseudo-randomly assigned as left leading images, 6 (including 3 electronics) served as
125 were appointed as right leading images, another 6 (including 3 electronics) served as right
126 trailing images while the remaining 6 (including 3 electronics) acted as right
127 trailing images. Therefore, each specific image could occur in any position or

128 condition (left or right, leading or trailing), thereby minimizing potential biases by
129 specific features of individual object stimuli. Image size was 5° x 5° visual angle
130 presented on a mid-gray background. Stimuli and their association remained the
131 same during the behavioral learning session, MRI scanning and a post-scanning
132 object categorization task. During the behavioral learning session and post-scanning
133 test, object stimuli were presented on an LCD screen (ASUS VG278q, 1920 x 1080
134 pixel resolution, 60 Hz refresh rate). During MRI scanning, stimuli were displayed on
135 a rear-projection MRI-compatible screen (SAMTEC SA-9900 projector, 1024 x 768
136 pixel resolution, 60 Hz refresh rate), visible using an adjustable mirror mounted on
137 the head coil.

138

139 **Experimental design**

140 Each participant completed two sessions on two consecutive days. The first session
141 comprised a behavioral learning task while the second session included an fMRI task
142 and a post-scanning object categorization task. While the stimuli and their
143 associations were identical during both sessions, different tasks were employed.

144 *Day one - Learning session.* Each trial began with a black fixation dot
145 (diameter = 0.4° visual angle) in the center of the screen, participants were asked to
146 maintain fixation on the fixation dot throughout the trial. Two leading images were
147 presented 1.5° visual angle left and right from the central fixation dot for 500 ms,
148 immediately followed by two trailing images, without ISI, at the same locations for
149 500 ms (**Figure 1a**). Participants were required to count the pairs of same category
150 objects (electronic vs. non-electronic) shown during the leading and trailing images
151 and respond within 2000 ms after trailing image onset by pressing one of three
152 response buttons (corresponding to none, one, or both; see *Pair counting task* below

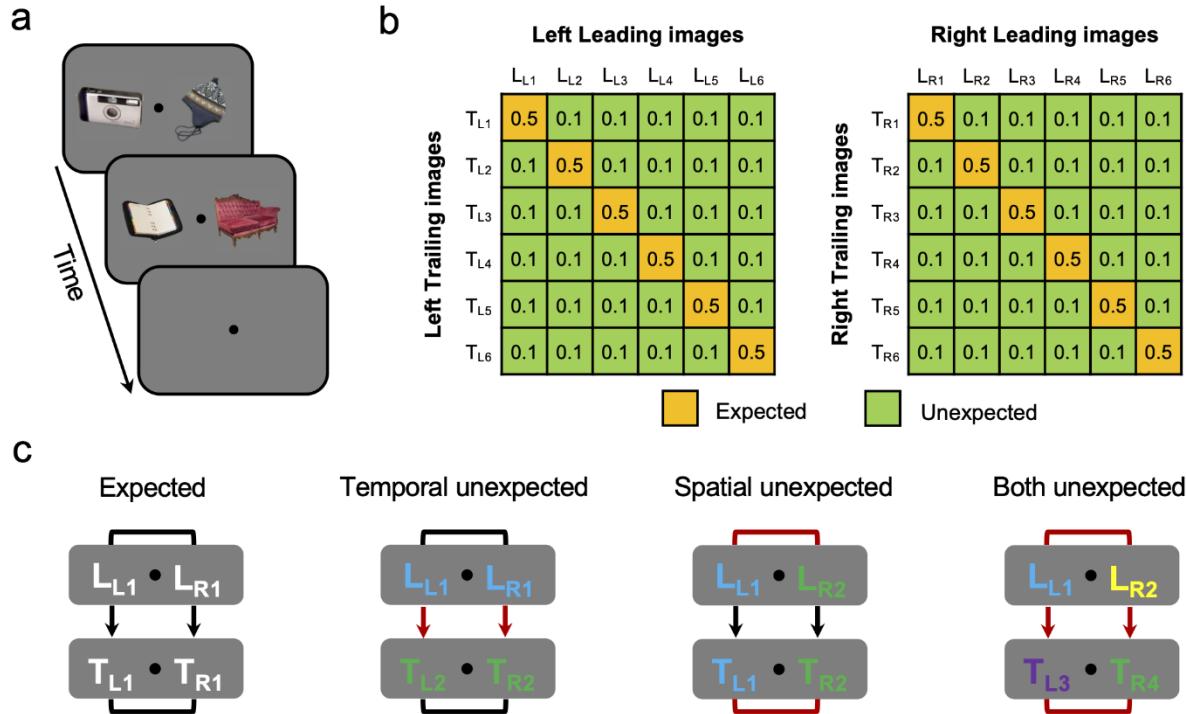
153 for details). Finally, feedback was presented for 500 ms, followed by a 1000 - 2000
154 ms ITI. 24 object images (12 electronics and 12 non-electronics) were pseudo-
155 randomly preselected per participant from a pool of images, 12 of which were
156 pseudo-randomly combined into pairs, forming a total of 6 leading image pairs (i.e.,
157 the first two images on a trial), while the remaining 6 pairs were used as trailing
158 image pairs (i.e., the second two images on a trial). Crucially, during the learning
159 session, the leading image pair was perfectly predictive of the identity of the trailing
160 image pair [$P(\text{trailing pair} | \text{leading pair}) = 1$]. At the same time, the left and right
161 images within both the leading and trailing image pairs were 100% predictive of one
162 another (i.e., pairs always occurred together). Thus resulting in deterministic
163 association in both spatial (co-occurrence) and temporal (sequence) contexts during
164 learning session (see the most left panel in **Figure 1c**). During the learning session
165 each participant performed 5 blocks, with each block comprised of 216 trials,
166 resulting in a total of 180 trials per pair during learning session. The learning session
167 took approximately 60 minutes.

168 *Day two – fMRI session.* One day after the learning session, participants
169 performed the fMRI session. This session started with one additional block identical
170 to the behavioral learning session, including 216 trials, to renew the learned
171 associations before MRI scanning. During MRI scanning, participants first performed
172 36 practice trials during acquisition of the anatomical image. The fMRI session was
173 similar to the behavioral learning session, except for the following three
174 modifications. First, a longer ITI of 2000 – 6000 ms (mean = 3000 ms) was used.
175 Second, instead of counting pairs of the same category, participants were required to
176 detect oddball images. Oddballs were the same object images, as shown before, but
177 flipped upside-down, occurring on 10% of trials. Participants were instructed to

178 respond to these target images by pressing a button as quickly as possible, while no
179 response was required during trials without an oddball image. Crucially, whether an
180 image was upside-down was completely randomized and could not be predicted on
181 the basis of the statistical regularities that were present in the image sequences.
182 Third, while the association between images remained the same as during the
183 behavioral learning session, in the fMRI session also unexpected image pairs were
184 shown. In particular, the transition matrices shown in **Figure 1b**, determined how
185 often images were presented together. In 50% of trials, a leading image pair was
186 followed by its expected trailing image pair, identical to the learning session, thus
187 constituting the expected condition. For instance, L_{L1} (leading image, left 1) and L_{R1}
188 (leading image, right 1) served as leading image pair for T_{L1} (trailing image, left 1)
189 and T_{R1} (trailing image, right 1). In the other half of trials, one of the three unexpected
190 conditions (temporally unexpected context, spatially unexpected context, both
191 temporally and spatially unexpected context) occurred with equal possibilities,
192 resulting in 16.67% per unexpected condition. Specifically, for the temporal
193 unexpected context (**Figure 1c** left middle panel), after presenting a leading image
194 pair, one of the other five unmatched trailing image pairs would occur. Thus, while
195 the two images within both the leading and trailing image pair were still expected
196 (i.e., no spatial expectation violation), the temporal sequence of images was
197 unexpected. For example, in this condition L_{L1} and L_{R1} were followed by T_{L2} and T_{R2} .
198 For the spatially unexpected context (**Figure 1c** right middle panel), each leading
199 image was followed by its expected trailing image (e.g., $L_{L1} \rightarrow T_{L1}$ and $L_{R2} \rightarrow T_{R2}$).
200 However, the two images presented during both the leading and trailing image
201 period were not usually paired; e.g., $L_{L1} \times L_{R2}$ and $T_{L1} \times T_{R2}$). Thus, in this condition
202 spatial context expectations were violated, while temporal context was expected,

203 thus constituting the spatially unexpected condition. In a final condition, both, spatial
204 and temporal context were violated (**Figure 1c** most right panel). In particular, all
205 four images shown during this condition did not appear together in the learning
206 session. Crucially, the expectation status only depended on the usual association
207 between the leading image pair and trailing image pair, rather than the frequency or
208 identity of an object image per se. In other words, each object image occurred as
209 expected object and in each unexpected condition. Therefore, all images occurred
210 equally often throughout the experiment, ruling out potential confounds of stimulus
211 frequency or familiarity. Feedback on behavioral performance (accuracy) was
212 provided after each run.

213 During MRI scanning, each run consisted of 108 trials, including 54 expected
214 trials, 18 temporal context violation trials, 18 spatial context violation trials and 18
215 trials where both spatial and temporal context were violated. The order of trials was
216 randomized within each run. In total each participant performed 5 runs. Each run
217 lasted ~12 minutes with 5 null events of 12 s that were evenly distributed across the
218 run, which also served as brief resting periods. The first 8 s of fixation was discarded
219 from analysis. Finally, after MRI scanning, a pair counting task, identical to the
220 learning session was performed outside of the MRI scanner room, which took
221 approximately 20 minutes (see *Pair counting task* below for details).



222

223 **Figure 1. Experimental paradigm and design. (a)** Experimental paradigm in both
224 the behavioral learning and fMRI session. A trial starts with a 500 ms presentation of
225 two leading images, presented left and right from the central fixation dot. The two
226 leading images are immediately followed by the trailing images, without ISI, at the
227 same locations, also shown for 500 ms. Participants were asked to detect an
228 infrequently presented upside-down version of the images (~10% of trials). Trials were
229 separated by a 2 - 6 s (mean 3 s) ITI period. **(b)** Shown are the image transition
230 matrices determining the statistical regularities between leading and trailing images
231 during MRI scanning. On the left, L_{L1} to L_{L6} represent the six leading images presented
232 on the left of the fixation dot, while T_{L1} to T_{L6} represent the associated six left trailing
233 images. Similarly, L_{R1} to L_{R6} represent the six right leading images, while T_{R1} to T_{R6}
234 represent the six right trailing images. Yellow cells indicate image pairs that are
235 expected by temporal context, while green denotes unexpected image pairs. Numbers
236 represent the probability of that cell during MRI scanning. Crucially, the left and right

237 images were also associated with each other, constituting the spatial context. For
238 instance, L_{L1} was associated with L_{R1} , and T_{L1} was associated with T_{R1} . In this case,
239 L_{L1} , L_{R1} , T_{L1} and T_{R1} composed two image pairs that were expected in both the
240 temporal and spatial contexts (see **Figure1c**, 'Expected'). **(c)** Illustration of the four
241 expectation conditions during MRI scanning. Black lines indicate expected
242 associations, while red lines indicate unexpected pairings. *Expected condition*: the
243 matched image configuration that was shown during the behavioral learning session.
244 *Temporally unexpected context*: both the two leading images (L_{L1} and L_{R1}) and two
245 trailing images (T_{L2} and T_{R2}) were expected in terms of spatial context (same as the
246 expected condition), the temporal association was violated (i.e., $L_{L1} \rightarrow T_{L2}$ and $L_{R1} \rightarrow$
247 T_{R2}). *Spatially unexpected context*: while the leading image reliably predicted the
248 identity of the trailing image on both the left ($L_{L1} \rightarrow T_{L1}$) and right ($L_{R2} \rightarrow T_{R2}$) side
249 independently, thus retaining the expected temporal context, image pairs were not
250 associated in terms of spatial context, neither during the leading images nor during
251 the two trailing images (e.g., L_{L1} and L_{R2} occurring together). *Both unexpected*: shown
252 were four images that do not appeared together in the expected condition. Therefore,
253 the expectation violations occurred in both the temporal and spatial contexts.

254 *Functional localizer*. Following the main task runs during the fMRI session,
255 two functional localizer runs were scanned. These localizer runs were used to define
256 object-selective LOC, and to select voxels that were maximally responsive to the
257 relevant object images. For each participant, the same 12 trailing images that were
258 previously seen in the main task runs and their phase-scrambled version were
259 presented during the localizer. Images were presented at the left and right from the
260 center of screen, corresponding to the location where the stimuli were shown during
261 the main task runs. Each image was shown for 11 s, alternating between the left and

262 right side. Images flashed with a frequency of 2 Hz (300 ms on, 200 ms off).
263 Throughout the localizer, participants were instructed to fixate the fixation dot, while
264 monitoring for an unpredictable dimming of the stimulus (dimming period = 300 ms).
265 Participants responded as quickly as possible by pressing a button. In each run, 4
266 null events of 11 s were evenly inserted, and each trailing image and its phase-
267 scrambled version was presented two times. The order of trials was fully
268 randomized, except for excluding direct repetitions of the same image. Each
269 participant completed two localizer runs, with each run lasting ~9.5 minutes. In total
270 each image and its phase-scrambled version was presented 4 times.

271 *Pair counting task.* Because the oddball detection performed during fMRI
272 scanning does not relate to the underlying statistical regularities, and therefore does
273 not indicate whether statistical regularities were indeed learned, an additional pair
274 counting task was performed after fMRI scanning. In this task, participants were
275 asked to count the number of pairs of the same object category shown on each trial.
276 Participants were further instructed to respond as quickly and accurately as possible.
277 Thus, this task was the same as the task performed during the behavioral learning
278 session, except that the three unexpected conditions were also included. The
279 rationale of this task was to gauge the learning of the object pairs (i.e., statistical
280 regularities) in terms of both temporal and spatial context. Participants could benefit
281 from the knowledge of the associations between the image pairs, as both knowledge
282 about the co-occurrence and temporal sequence would allow for faster responses.
283 Therefore, the performance difference (e.g., accuracy and reaction time) between
284 the expected condition and each unexpected condition could be considered as an
285 indication for having learnt the underlying statistical regularities. In total, participants
286 performed 360 trials split into 2 blocks, including 180 expected trials, 60 temporally

287 unexpected context trials, 60 spatially unexpected context trials and 60 trials in which
288 both spatial and temporal context were unexpected. The pair counting task took
289 approximately 20 minutes.

290 **fMRI parameters**

291 Functional and anatomical images were acquired on a 3.0T GE MRI-750 system (GE
292 Medical Systems, Waukesha, WI, USA) at Hangzhou Normal University, using a
293 standard 8-channel headcoil. Functional images were acquired in a sequential
294 (ascending) order using a T2*-weighted gradient-echo EPI pulse sequence (TR/TE =
295 2000/30 ms, voxel size $2.5 \times 2.5 \times 2.3$ mm, 0.2 mm slice space, 36 transversal
296 slices, 75° flip angle, FOV = 240 mm²). Anatomical images were acquired using a
297 T1-weighted inversion prepared 3D spoiled gradient echo sequence (IR-SPGR)
298 (inversion time = 450 ms, TR/TE = 8.2/3.1ms, FOV = 256 × 256 mm², voxel size 1 ×
299 1 × 1 mm, 176 transversal slices, 8° flip angle, parallel acceleration = 2).

300 **Data analysis**

301 ***Behavioral data analysis***

302 Behavioral data from the pair counting task was analyzed in terms of response
303 accuracy and RT. RT was calculated relative to the onset of the trailing image
304 objects. Only trials with correct responses were included in RT analysis. Additionally,
305 we excluded trials with RTs shorter than 200 ms (0.82%) or more than three
306 standard deviations above the subject's mean response time (0.49%). RT and
307 accuracy data for expected and unexpected trailing image trials were averaged
308 separately per participant and across subjects subjected to a paired t test. The effect
309 size was calculated in terms of Cohen's d_z for all paired t-test, while partial eta-
310 squared (η^2) was used for indicating effect sizes in the repeated measures ANOVA
311 (Lakens, 2013).

312 ***fMRI data preprocessing***

313 fMRI data preprocessing was performed using FSL 6.0.1 (FMRIB Software Library;
314 Oxford, UK; www.fmrib.ox.ac.uk/fsl; Smith et al., 2004, RRID:SCR_002823). The
315 preprocessing pipeline included brain extraction (BET), motion correction
316 (MCFLIRT), slice timing correction (Regular up), temporal high-pass filtering (128 s),
317 and spatial smoothing for univariate analyses (Gaussian kernel with FWHM of 5
318 mm). Functional images were registered to the anatomical image using FSL FLIRT
319 (BBR) and to the MNI152 T1 2 mm template brain (linear registration with 12
320 degrees of freedom). Registration to the MNI152 template brain was only applied for
321 whole-brain analyses, while all ROI analyses were performed in each participant's
322 native space in order to minimize data interpolation.

323 ***Whole brain analysis***

324 To estimate the BOLD response to expected and unexpected stimuli across the
325 entire brain, FSL FEAT was used to fit voxel-wise general linear models (GLM) to
326 each participant's run data in an event-related approach. In the first level GLMs,
327 expected and three unexpected image object trials were modeled as four separate
328 regressors with a duration of one second (the combined duration of leading and
329 trailing image pairs), and convolved with a double gamma hemodynamic response
330 function. An additional nuisance regressor for oddball trials (upside-down images)
331 was added. Additionally, first-order temporal derivatives for the five regressors, and
332 24 motion regressors (FSL's standard + extended motion parameters) were also
333 added to the GLM. To quantify the main effects of spatial and temporal expectation
334 suppression, we contrasted unexpected regressors and the expected regressors for
335 spatial and temporal context separately (i.e., temporal context expectation
336 suppression = $BOLD_{Temporal\ unexpected} + BOLD_{Both\ unexpected} - BOLD_{Spatial\ unexpected} -$

337 $BOLD_{\text{Both expected; spatial context expectation suppression}} = BOLD_{\text{Spatial unexpected}} +$
338 $BOLD_{\text{Both unexpected}} - BOLD_{\text{Temporal unexpected}} - BOLD_{\text{Both expected}}).$ Data were combined
339 across runs using FSL's fixed effect analysis. For the across-participants whole-brain
340 analysis, FSL's mixed effect model (FLAME 1) was used. Multiple-comparison
341 correction was performed using Gaussian random-field based cluster thresholding.
342 The significance level was set at a cluster-forming threshold of $z > 3.1$ (i.e., $p <$
343 0.001 , two-sided) and a cluster significance threshold of $p < 0.05$.

344 ***Regions of interest (ROIs) analysis***

345 ROI analyses were conducted in each participant's native space. Primary visual
346 cortex (V1), object-selective lateral occipital complex (LOC), and temporal occipital
347 fusiform cortex (TOFC) were chosen as the three ROIs (see *ROI definition* below) for
348 analysis, based on two previous studies that used a similar experimental design
349 (Richter et al., 2018; Richter and de Lange, 2019). The mean parameter estimates
350 were extracted from each ROI for the expected and unexpected conditions
351 separately. For each ROI, these data were submitted to a two-way repeated
352 measures ANOVA with temporal context (expected vs. unexpected) and spatial
353 context (expected vs. unexpected) as factors.

354 ***ROI definition.*** All ROIs were defined using independent data from the localizer
355 runs. Specifically, V1 was defined based on each participant's anatomical image,
356 using Freesurfer 6.0 to define the gray–white matter boundary and perform cortical
357 surface reconstruction (recon-all; Dale et al., 1999; RRID:SCR_001847). The
358 resulting surface-based ROI of V1 was then transformed into the participant's native
359 space and merged into one bilateral mask. Object selective LOC was defined as
360 bilateral clusters, within anatomical LOC, showing a significant preference for intact
361 compared to scrambled object stimuli during the localizer run (Kourtzi and

362 Kanwisher, 2001; Haushofer et al., 2008). To achieve this, intact objects and
363 scrambled objects were modeled as two separate regressors in each participant's
364 localizer data. The temporal derivatives of all regressors and the 24 motion
365 regressors were also added to fit the data. Finally, the contrast of interest, objects
366 minus scrambles, was constrained to anatomical LOC. In order to create the TOFC
367 ROI mask, the anatomical temporal-occipital fusiform cortex mask from the Harvard-
368 Oxford cortical atlas (RRID:SCR_001476), distributed with FSL, was further
369 constrained to voxels showing a significant conjunction inference of expectation
370 suppression on the group level in Richter et al. (2018) and Richter and de Lange
371 (2019). The resulting mask was then transformed from MNI space to each
372 participant's native space using FSL FLIRT. Finally, the 200 most active voxels in
373 each of the three ROI masks were selected for further statistical analyses. To this
374 end, the contrast interest between the left and right hemisphere in V1 (including both
375 the intact and scrambled images) was calculated, while in LOC and TOFC, the
376 contrast interest between the intact images and the scrambled images was
377 calculated based on the localizer data. The resulting z-map of this contrast was then
378 averaged across runs. Finally, we selected the 200 most responsive voxel from this
379 contrast. In order to verify that our results did not depend on the a priori defined, but
380 arbitrary number of voxels in the ROI masks, we repeated all ROI analyses with
381 masks ranging from 50 to 500 voxels in steps of 50 voxels.

382 ***Bayesian analysis***

383 In order to further evaluate any non-significant results, and arbitrate between an
384 absence of evidence and evidence for the absence of an effect, the Bayesian
385 equivalents of the above outlined analyses were additionally performed. JASP 0.10.2
386 (JASP Team, 2019, RRID:SCR_015823) was used to perform all Bayesian analyses,

387 using default settings. Thus, for Bayesian t-tests a Cauchy prior width of 0.707 was
388 chosen. Qualitative interpretations of Bayes Factors are based on criteria by Lee and
389 Wagenmakers (2014).

390

391

392 **Results**

393 We exposed participants to statistical regularities by presenting two
394 successive object image pairs in which the leading image pairs predicted the identity
395 of the trailing image pairs. The identities of the image pairs were also predictable in
396 terms of their spatial context; i.e., simultaneously shown left and right images
397 occurred together. Subsequently, in the MRI scanner, participants were shown the
398 same predictable object image pairs (expected condition), but additional expectation
399 violations were introduced. In particular, either the temporal context was violated, the
400 spatial context was violated, or both contexts were violated (see **Figure 1c**).

401

402 **Stronger modulation of spatial context than temporal context on sensory
403 processing throughout the ventral visual stream**

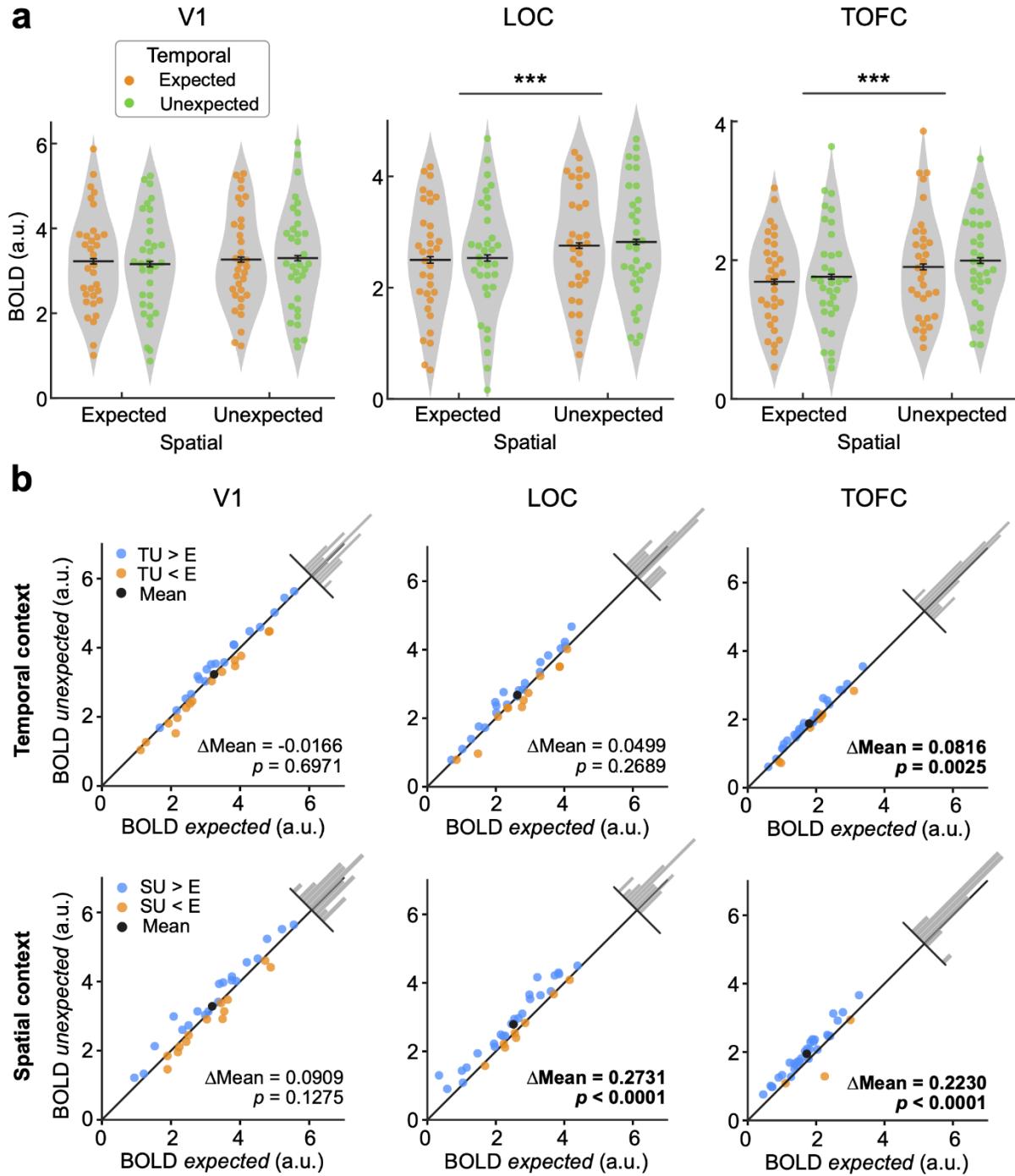
404 In order to assess the consequences of violating temporal and spatial context
405 expectations we performed a two-way repeated measures ANOVA with temporal
406 context (expected vs. unexpected) and spatial context (expected vs. unexpected) as
407 factors, within our a prior defined ROIs: primary visual cortex (V1), object-selective
408 lateral occipital complex (LOC), and temporal occipital fusiform cortex (TOFC). In
409 higher visual areas, LOC and TOFC, we observed a significant decrease in BOLD
410 responses when stimuli were expected in terms of their spatial context (**Figure 2a**;
411 LOC: $F_{(1, 32)} = 31.389, p = 3.0\text{e-}6, \eta^2 = 0.495$; TOFC: $F_{(1, 32)} = 23.083, p = 3.5\text{e-}5, \eta^2$

412 = 0.419). In other words, when two stimuli frequently co-occurred, thus making them
413 expected in this pair, they elicited reduced sensory responses in ventral visual areas.
414 Furthermore, we found a similar suppression of neural responses by temporal
415 context expectations in TOFC ($F_{(1, 32)} = 10.805, p = 0.0025, \eta^2 = 0.252$), but not in
416 LOC ($F_{(1, 32)} = 1.266, p = 0.2689, \eta^2 = 0.038$). That is, in TOFC, if a pair of stimuli
417 was expected given the preceding stimulus pair, the elicited BOLD response was
418 suppressed compared to the response to the same pair occurring in an unexpected
419 temporal sequence. No interaction between temporal and spatial context was found
420 in either LOC or TOFC (LOC: $F_{(1, 32)} = 0.111, p = 0.7412, \eta^2 = 0.003$; TOFC: $F_{(1, 32)} =$
421 $0.064, p = 0.8013, \eta^2 = 0.002$). Thus, the suppression of neural responses induced
422 by temporal expectations was not modulated by spatial context expectations, and
423 vice versa.

424 In a post-hoc analysis we compared the magnitude of neural suppression
425 induced by temporal and spatial context predictions. In LOC and TOFC spatial
426 context expectations resulted in a larger suppression than temporal expectations
427 (LOC: $t_{(32)} = 2.870, p = 0.0072$, Cohen's $d_z = 0.835$; TOFC: $t_{(32)} = 2.575, p = 0.0149$,
428 Cohen's $d_z = 0.691$), thus suggesting that spatial context may be a stronger
429 modulator of visual responses than temporal context.

430 Perhaps surprisingly, we did not find any reliable modulation of neural
431 responses by temporal or spatial context predictions in V1 (spatial context: $F_{(1, 32)} =$
432 $2.448, p = 0.1275, \eta^2 = 0.071$; temporal context: $F_{(1, 32)} = 0.154, p = 0.6971, \eta^2 =$
433 0.005 ; spatial context by temporal context interaction: $F_{(1, 32)} = 0.627, p = 0.4342, \eta^2 =$
434 0.019). Indeed, in V1, Bayesian analyses yielded moderate evidence for the
435 absence of a modulation of neural responses by temporal context violations
436 (temporally unexpected context vs. expected context: $BF_{10} = 0.141$), and anecdotal

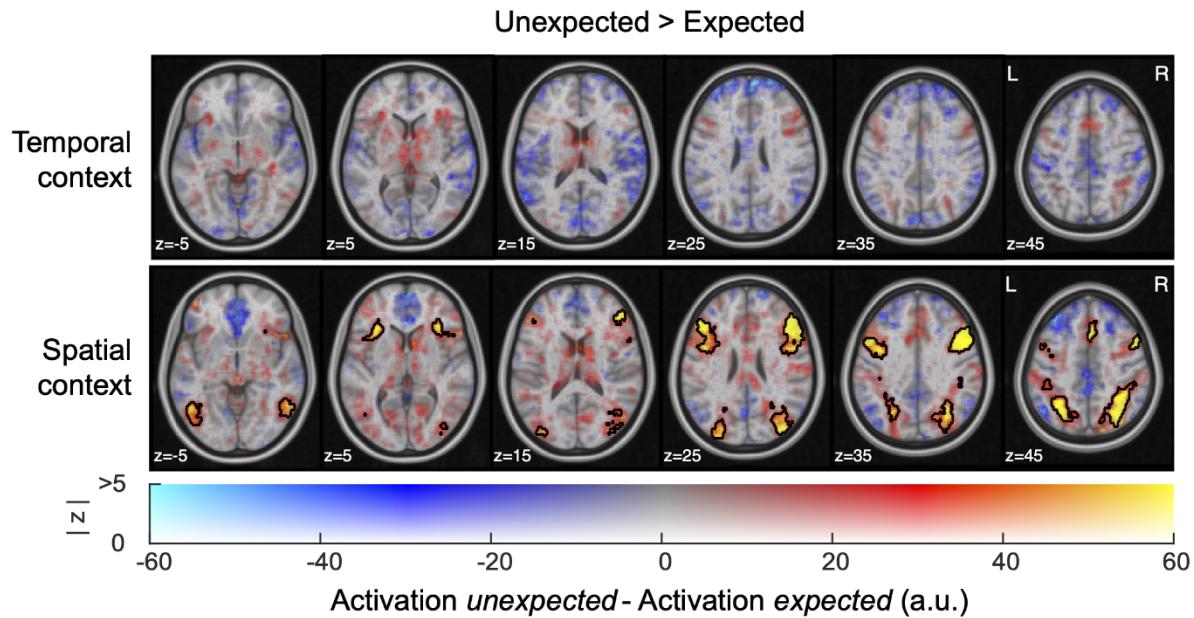
437 support for the absent of an effect when spatial context was violated (spatially
438 unexpected context vs. expected context: $BF_{10} = 0.388$). Thus, in V1 expectations, in
439 terms of temporal or spatial context, did not appear to modulate sensory responses.
440 In contrast, in higher visual areas a suppression of responses to expected stimuli
441 was observed both for temporal and spatial contexts.



442 **Figure 2.** Expectation suppression within V1, LOC and TOFC. **(a)** Parameter
 443 estimates for responses to expected and unexpected images pairs. In both LOC and
 444 TOFC, BOLD responses to spatially expected image pairs were significantly
 445 attenuated compared to unexpected image pairs. Furthermore, a reliable suppression
 446 of responses by temporal context expectations was observed in TOFC. No modulation
 447 of BOLD responses by expectations was found in V1. Each dot denotes an individual
 448

449 participant and the black line is the mean across participants. Error bars denote ± 1
450 within-subject SEM. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$. **(b)** BOLD responses evoked
451 by unexpected and expected context within V1 (left column), LOC (middle column)
452 and TOFC (right column). The upper row represents the BOLD contrast between the
453 temporally unexpected context and expected context, averaged across the spatially
454 expected and unexpected context. The bottom row represents the BOLD contrast
455 between the spatially unexpected context and expected context, averaged across the
456 temporally expected and unexpected context. Blue and yellow dots represent
457 individual participants. Blue indicates expectation suppression (unexpected >
458 expected), yellow indicates expectation enhancement (unexpected < expected), and
459 black indicates the mean of all subjects. Δ Mean is equal to the difference of BOLD
460 response between the unexpected and expected condition. The inset histogram
461 shows the distribution of deviations from the unity line.

462 To ensure that our results were not dependent on the a prior but arbitrarily
463 chosen mask sizes of the ROIs, we repeated the analyses for ROIs of sizes ranging
464 from 50 to 500 voxels in step of 50 voxels. Results were qualitatively identical to
465 those mentioned above (**Figure 2a**) for all ROI sizes within all three ROIs (V1, LOC,
466 TOFC), indicating that our results do not depend on ROI size, but well represent
467 results within the ROIs.



468

469 **Figure 3.** Expectation suppression across cortex for temporal and spatial contexts.
470 Displayed are parameter estimates for unexpected minus expected image pairs
471 overlaid onto the MNI152 2 mm anatomical template. Color represents the
472 unthresholded parameter estimates: red-yellow clusters denote expectation
473 suppression, blue-cyan clusters indicate expectation enhancement; opacity indicates
474 the z statistics of the contrast. Black contours outline statistically significant clusters
475 (Gaussian random field cluster corrected). No significant clusters were found for the
476 main effect of temporal context (upper row). The main effect of spatial expectation
477 (bottom row) shows significant clusters of expectation suppression in parts of the
478 ventral visual stream (LOC, TOFC), as well as bilateral frontal gyrus, bilateral
479 precentral gyrus, bilateral frontal operculum and insular cortex, and paracingulate
480 gyrus.

481 A complementary whole-brain analysis was performed to investigate the effect
482 of temporal context and spatial context outside of our predefined ROIs. Results are
483 illustrated in **Figure 3**. In accordance with our ROI analysis, spatial expectations

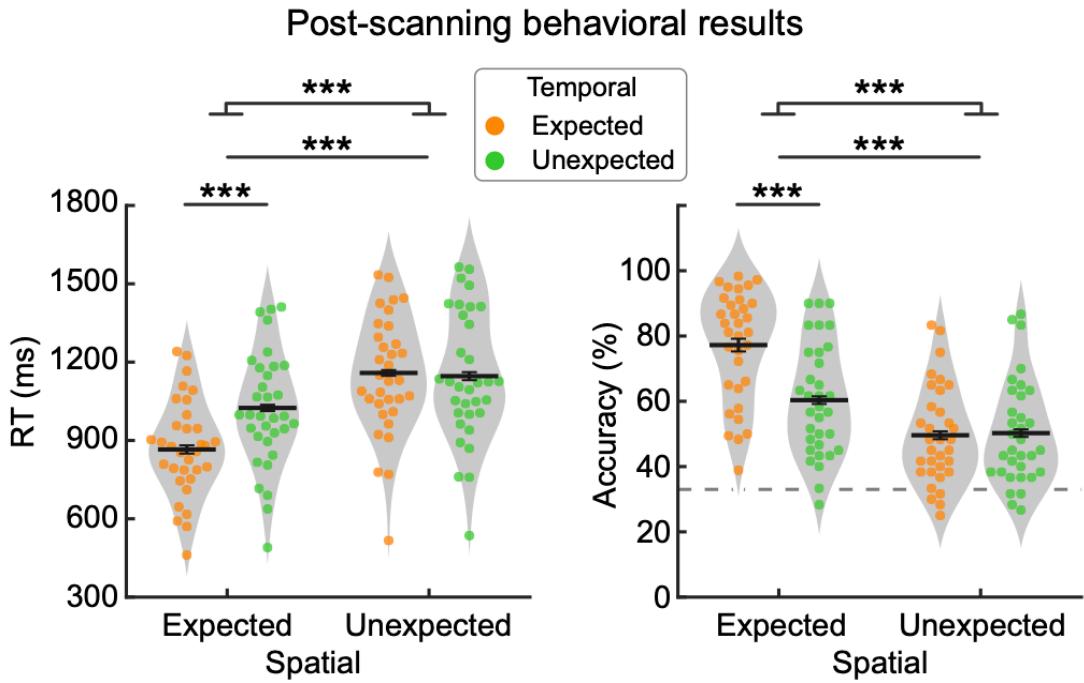
484 were associated with significantly suppressed neural responses throughout the
485 ventral visual stream. Additional clusters of expectation suppression were evident
486 outside the ventral visual stream, including bilateral frontal gyrus, bilateral precentral
487 gyrus, bilateral frontal operculum and insular cortex, as well as the paracingulate
488 gyrus. In contrast, no reliable modulation by temporal context expectation was found
489 outside of our predefined ROIs in the whole-brain analysis. Thus, temporal context
490 expectations were only evident in the ROI analysis, but too small or hidden by
491 interindividual variability to be detected in the whole-brain analysis (note: ROI masks
492 were individually defined for each participant; also see Materials and Methods, ROI
493 definition).

494 **Expectations facilitate object categorization**

495 In addition to the neural effects of expectations, we also examined whether
496 expectations facilitated behavioral responses. During a post-scanning object
497 categorization task, participants were asked to count the number of object pairs of
498 the same category shown as leading and trailing image pairs (i.e., 0, 1 or 2 pairs
499 could be of the same category). In order to fulfill this task, as quickly and accurately
500 as possible, participants could benefit from the knowledge of the underlying
501 statistical regularities – both in terms of co-occurrence (spatial) and sequence
502 (temporal) prediction. In line with our hypothesis, RTs and accuracy of responses
503 (**Figure 4**) were affected by expectations, in both temporal (RT: $t_{(32)} = 4.891$, $p =$
504 $6.9\text{e-}6$, Cohen's $d_z = 0.851$; accuracy: $t_{(32)} = 4.924$, $p = 6.1\text{e-}6$, Cohen's $d_z = 0.857$)
505 and spatial contexts (RT: $t_{(32)} = 11.670$, $p = 1.3\text{e-}17$, Cohen's $d_z = 2.031$; accuracy:
506 $t_{(32)} = 10.224$, $p = 3.7\text{e-}15$, Cohen's $d_z = 1.780$). Thus, participants learned and
507 benefitted from both spatial and temporal context predictions.

508 Interestingly, participants were faster and more accurate in response to
509 objects predicted by the temporal sequence only when the spatial context was
510 expected as well (RT: $t_{(32)} = 9.329$, $p = 1.2\text{e-}10$, Cohen's $d_z = 1.624$; accuracy: $t_{(32)} =$
511 7.649 , $p = 1.0\text{e-}8$, Cohen's $d_z = 1.332$), but not when the spatial context was
512 unexpected (RT: $t_{(32)} = 0.269$, $p = 0.7898$, Cohen's $d_z = 0.047$, $\text{BF}_{10} = 0.193$;
513 accuracy: $t_{(32)} = 0.566$, $p = 0.5755$, Cohen's $d_z = 0.099$, $\text{BF}_{10} = 0.216$). The
514 robustness of this distinct pattern of facilitation effect was statistically confirmed by
515 an interaction analysis (RT: $F_{(1, 32)} = 38.787$, $p = 5.6\text{e-}7$, $\eta^2 = 0.548$; accuracy: $F_{(1, 32)}$
516 $= 46.337$, $p = 1.1\text{e-}7$, $\eta^2 = 0.592$). Moreover, when a stimulus was expected by
517 spatial context, participants showed faster and more accurate responses,
518 irrespective of whether the temporal context was expected (RT: $t_{(32)} = 13.977$, $p =$
519 $3.6\text{e-}15$, Cohen's $d_z = 2.433$; accuracy: $t_{(32)} = 10.883$, $p = 2.7\text{e-}12$, Cohen's $d_z =$
520 1.894) or unexpected (RT: $t_{(32)} = 5.838$, $p = 1.7\text{e-}6$, Cohen's $d_z = 1.016$; accuracy:
521 $t_{(32)} = 6.279$, $p = 4.9\text{e-}7$, Cohen's $d_z = 1.093$).

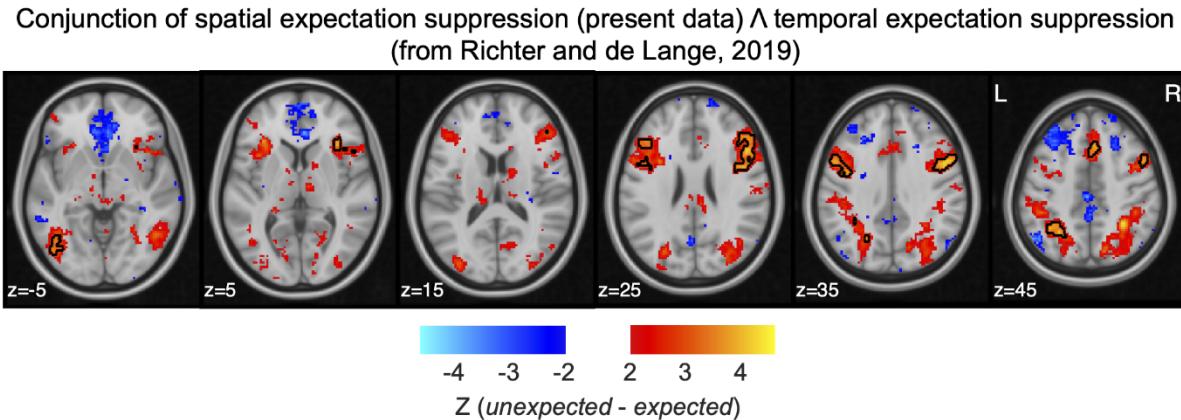
522 In sum, behavioral performance was reliably facilitated by spatial context,
523 resulting in faster and more accurate responses. On the other hand, expected
524 temporal sequences also aided in faster and more accurate responses, however
525 only when the spatial context was expected. These results may suggest that
526 participants grouped pairs of objects, and predicted the upcoming pair of objects,
527 instead of individual sequences of objects on the left and right side separately.



528

529 **Figure 4.** Behavioral data indicate statistical learning. Reaction time (left) and
530 accuracy (right) are plotted for expected and unexpected conditions in temporal (dot
531 color) and spatial contexts (abscissa), respectively. Behavioral responses in the
532 spatially expected condition are significantly faster and more accurate than in the
533 unexpected condition. Temporally expected stimulus pairs also result in faster and
534 more accurate responses, however this effect is only present when spatial
535 expectations were met. Dashed horizontal gray line indicates chance level accuracy
536 (33.33%). Dots represent single subject data. Black line is the mean across
537 participants. Error bars denote ± 1 within-subject SEM. *** $p < 0.001$.

538 **Spatial and temporal context expectations modulate neural responses in
539 similar cortical areas**



540
541 **Figure 5** Displayed are z statistics of the contrast between unexpected and expected
542 of a conjunction inference between data from the spatial context violation and data
543 from a temporal context violation effect from Richter and de Lange (2019). Red-yellow
544 clusters denote expectation suppression. Significant overlaps in the localization of
545 expectation suppression include clusters in parts of the ventral visual stream, middle
546 and inferior frontal gyrus, and precentral gyrus.

547 Given the modulation of neural responses by temporal context in TOFC in our
548 ROI analysis (**Figure 2a**), and the reliability of expectation suppression reported in
549 previous studies investigating temporal context violations (Turk-Browne et al., 2009;
550 Meyer and Olson, 2011; Richter and de Lange, 2019), it is perhaps surprising that
551 we did not find evidence of temporal expectation effects in the whole-brain analysis
552 (**Figure 3**). Potential explanations why temporal context violations did show little
553 effect in the present study will be discussed in more detail later (see Discussion).
554 However, in order to further compare spatial and temporal predictions, it could be
555 informative to compare the localization of the here reported spatial expectation
556 suppression with temporal expectation suppression shown in previous studies. In a
557 conjunction analysis, we investigated the overlap of expectation suppression
558 between previously reported temporal expectation suppression from Richter and de
559 Lange (2019) and the present spatial expectation violation. Results illustrated in

560 **Figure 5**, show clusters of overlapping expectation suppression between temporal
561 and spatial context expectations throughout parts of the ventral visual stream, and
562 several non-sensory areas, including middle and inferior frontal gyrus, precentral
563 gyrus. Thus, spatial context expectations, as observed here, and temporal context
564 expectations, as reported by Richter and de Lange (2019), are evident in a similar
565 neural network, thereby suggesting that a comparable neural mechanism may
566 underlie both spatial and temporal context predictions.

567

568 **Discussion**

569 Both spatial and temporal context play an important role in visual perception
570 and behavior (Schwartz et al., 2007). The present study investigated the neural
571 consequences of violations of expectations derived from spatial and temporal
572 context, across the ventral visual stream. To this end, we exposed participants to two
573 forms of statistical regularities, making stimuli predictable in terms of spatial context
574 (co-occurrence of stimuli at specific locations) and temporal context (specific
575 temporal sequence of stimuli). While we measured brain activity to these stimuli,
576 image transitions were not task relevant, and thus any neural modulations by spatial
577 and temporal context were not dependent on task-relevance of the underlying
578 statistical regularities. We found a reliable and wide-spread activity modulation in the
579 ventral visual stream, including LOC and TOFC, as a function of spatial context. In
580 particular, when stimuli frequently co-occurred neural responses were suppressed
581 compared to the response to the same stimulus co-occurring with another stimulus,
582 even though all stimuli were equally familiar and always occurred at the same spatial
583 location. Temporal context (i.e., predictability of stimulus sequence) also modulated

584 neural responses in TOFC, again evident as a suppression of responses to expected
585 stimuli. Interestingly, while the two forms of context modulated overlapping regions,
586 the activity modulation by spatial context was much stronger and more wide-spread
587 than the modulation by temporal context. Thereby our results extend previous
588 studies (e.g., Summerfield et al., 2008; Alink et al., 2010; Kok et al., 2012; Richter et
589 al., 2018; Richter and de Lange, 2019) by demonstrating that spatial and temporal
590 context priors may modulate neural responses in a similar fashion and within the
591 same cortical network. However, at least in the visual system spatial context appears
592 to be a more potent modulator of perceptual processing than temporal context.

593

594 **Spatial and temporal context facilitate behavior**

595 Our data showed a substantial and robust facilitation of behavioral responses
596 by both spatial and temporal contexts. During a post-scanning test, requiring
597 participants to count stimulus pairs of the same category (i.e., both electronic, or
598 both non-electronic stimuli), spatial and temporal context strongly modulated
599 behavioral performance (**Figure 4**). Specifically, responses were faster and more
600 accurately to stimuli presented in a spatially and temporally expected context, and
601 the violation of either context increased RTs and decreased response accuracy –
602 with larger decrements for spatial context violations. Crucially, the benefit of
603 temporally expected contexts was only observed when the spatial context was
604 expected. However, performance enhanced by spatially expected contexts was
605 evident irrespective of whether temporal context expectations were confirmed or
606 violated.

607 Thus, our data show that participants can in principle learn and benefit from
608 both spatial and temporal statistical regularities. However, our results also suggest
609 that our participants may have grouped simultaneously presented objects into image
610 pairs, which combined predicted the next image pair. That is, even though object
611 stimuli on the left and right side predicted the identity of the next stimulus
612 independently, even when spatial configuration were unexpected, these statistical
613 regularities may not have been learned, or the resulting predictions may not have
614 been instantiated. These results may suggest a preference for spatial over temporal
615 grouping in vision. However, it is important to note here that a strategy of grouping
616 spatial pairs may have partially been induced by the same-different category
617 counting task during learning, which specifically requires participants to make a
618 judgment about the groups of objects.

619

620 **Spatial and temporal context modulate sensory processing in the ventral**
621 **visual stream**

622 Our fMRI results show that sensory responses in object selective visual areas
623 (LOC and TOFC) are suppressed, if stimuli occur in expected spatial contexts
624 compared to unexpected spatial contexts. In other words, stimuli that frequently co-
625 occur evoked reduced sensory responses relative to the same stimuli presented in
626 less frequently co-occurring configurations. Note, that the frequency of the individual
627 stimuli occurring were equal, thereby excluding potentially confounding effects of
628 stimulus frequency or familiarity. Moreover, during MRI scanning predictions were
629 task-irrelevant, thus suggesting that predictions were formed and modulated neural
630 responses automatically.

631 The suppression of neural responses by spatial predictions matches key
632 characteristics of expectation suppression, a phenomenon previously described in
633 terms of suppressed sensory responses to stimuli expected by virtue of their
634 temporal context; i.e., a leading image predicting the identity of a trailing image (den
635 Ouden et al., 2009; Meyer and Olson, 2011; Richter et al., 2018; Richter and de
636 Lange, 2019). In line with previous studies, we also found a suppression of sensory
637 responses by temporal context in TOFC. That is, stimuli in expected temporal
638 sequences elicited suppressed BOLD responses compared to stimuli in unexpected
639 temporal sequences.

640 Moreover, using a conjunction analysis we showed that the here observed
641 spatial context suppression is evident in similar cortical areas as previously reported
642 suppression by temporal context expectations (e.g., den Ouden et al., 2009; Turk-
643 Browne et al., 2009, 2010; Gheysen et al., 2011; Meyer and Olson, 2011; Richter et
644 al., 2018; Richter and de Lange, 2019). Interestingly, this overlap in cortical regions
645 was not limited to object selective visual cortex, but also included several non-
646 sensory areas, such as inferior frontal gyrus. Combined these results suggest that
647 spatial and temporal contexts can have similar modulatory effects on neural
648 processing, thereby implying that the neural mechanism underlying contextual
649 prediction effects may be independent of the type of prediction – temporal or spatial
650 contexts. In agreement with this suggestion, Karuza et al. (2017) reported similar
651 neural modulations, and comparable correlations of these modulations with behavior,
652 during learning of spatial regularities as previously reported for statistical learning of
653 temporal (sequence) regularities (e.g., Turk-Browne et al., 2009, 2010; Gheysen et
654 al., 2010, 2011; Schapiro et al., 2014). Thus, the available data suggest that the
655 neural architecture and computations underlying different types of context

656 predictions may largely overlap, evident in similar modulations of both behavioral
657 and neural responses.

658

659 **Stronger modulations of neural responses by spatial context than temporal**
660 **context**

661 While the present data showed a joint modulation of neural responses by
662 spatial and temporal context, the modulation by temporal context was relatively
663 modest and significantly smaller than the modulation by spatial context. Initially,
664 these results may be surprising given the multitude of previous studies reporting
665 strong and extensive modulations of sensory responses by temporal context
666 predictions across the ventral visual stream (Turk-Browne et al., 2009, 2010;
667 Gheysen et al., 2010; Meyer and Olson, 2011; Tobia et al., 2012a, 2012b; Tremblay
668 et al., 2013; Plante et al., 2015; Richter et al., 2018; Richter and de Lange, 2019).
669 These previous studies however lacked spatial context, presenting single stimuli in
670 isolation.

671 Vision is particularly apt to handle simultaneous inputs and the spatial
672 structure between these stimuli (Saffran, 2002). Audition on the other hand shows a
673 remarkable sensitivity to the temporal structure of inputs (Kubovy, 1988; Conway
674 and Christiansen, 2009). Indeed, such modality specific constraints can affect the
675 manner in which stimuli are processed (Mahar et al., 1994; Repp and Penel, 2002),
676 maintained in working memory (Penney, 1989; Collier and Logan, 2000) and learned
677 (Handel and Buffardi, 1969; Saffran, 2002; Conway and Christiansen, 2009). Thus,
678 modality specific biases in the visual system may result in an emphasis on spatial

679 configurations and hence a stronger modulation of neural responses by spatial than
680 temporal context predictions.

681 Our behavioral results also support the notion that spatial predictions were
682 more readily acquired and utilized than temporal predictions. In particular, only when
683 spatial configurations were expected temporal predictions facilitated behavioral
684 responses. Thus, in the present data, and possibly vision in general, spatial
685 regularities appear to take precedence over temporal statistical regularities, resulting
686 in a larger magnitude of behavioral and neural modulations by spatial compared to
687 temporal context.

688

689 **No modulation of neural responses by prediction in primary visual cortex**

690 Surprisingly, we found no modulation by predictions in V1, unlike in some
691 previous studies (e.g., Kok et al., 2012; Richter and de Lange, 2019). It is possible
692 that, because expectations constitute a top-down modulation, likely originating from
693 beyond visual cortex (Hindy et al., 2019), its effect might be less pronounced in V1
694 compared to higher visual areas. Indeed, in previous studies prediction effects
695 appear to reduce in magnitude in lower visual areas (e.g. see Figure 1A in Richter
696 and de Lange, 2019). Moreover, it is possible that spatial arrangements of object
697 stimuli were too complex to yield specific predictions relevant to the response
698 properties of neural assemblies in V1. That is, predictions in our study constitute
699 arrangements and sequences of full color object images, thus particularly depending
700 on object selective cortical areas. Hence, arrangements of stimuli exploiting the
701 neural tuning in V1, such pairs of oriented grating stimuli may result in prediction
702 induced modulations in V1. Thus, the absence of expectation suppression in V1

703 observed here may be a consequence of the utilized stimuli and experimental
704 design.

705

706 **Conclusion**

707 In conclusion, our data suggest that temporal and spatial statistical
708 regularities jointly facilitate behavioral responses, leading to faster and more
709 accurate responses. At the same time, predictions based on both forms of contexts
710 modulate sensory responses, resulting in a suppression of responses to expected
711 stimuli in a similar cortical network, including object selective visual cortex. However,
712 spatial context appears a more potent modulator within the visual system, resulting
713 in larger modulations of neural responses by spatial compared to temporal context.

714

715 **Author contributions**

716 T.H., D.R., Z.W., and F.P.d.L. designed research; T.H. performed research; T.H. and
717 D.R. analyzed data; T.H. and D.R. wrote the first draft of the paper; T.H., D.R., Z.W.,
718 and F.P.d.L. edited the paper.

719

720 **References**

721 Alink A, Schwiedrzik C, Kohler A, Singer W, Muckli L (2010) Stimulus Predictability
722 Reduces Responses in Primary Visual Cortex. *The Journal of Neuroscience*
723 30:2960–2966.

724 Bar M (2004) Visual objects in context. *Nature Reviews Neuroscience* 5:617–629.

725 Bertels J, Franco A, Destrebecqz A (2012) How implicit is visual statistical learning?
726 *Journal of Experimental Psychology: Learning, Memory, and Cognition*
727 38:1425–1431.

728 Brady TF, Konkle T, Alvarez GA, Oliva A (2008) Visual long-term memory has a
729 massive storage capacity for object details. *PNAS* 105:14325–14329.

730 Collier GL, Logan G (2000) Modality differences in short-term memory for rhythms.
731 *Mem Cogn* 28:529–538.

732 Conway CM, Christiansen MH (2009) Seeing and hearing in space and time: Effects
733 of modality and presentation rate on implicit statistical learning. *European*
734 *Journal of Cognitive Psychology* 21:561–580.

735 Dale AM, Fischl B, Sereno MI (1999) Cortical Surface-Based Analysis: I.
736 Segmentation and Surface Reconstruction. *NeuroImage* 9:179–194.

737 Davenport JL, Potter MC (2004) Scene Consistency in Object and Background
738 Perception. *Psychol Sci* 15:559–564.

739 den Ouden HEM, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009) A Dual
740 Role for Prediction Error in Associative Learning. *Cereb Cortex* 19:1175–
741 1185.

742 Egner T, Monti JM, Summerfield C (2010) Expectation and Surprise Determine
743 Neural Population Responses in the Ventral Visual Stream. *J Neurosci*
744 30:16601–16608.

745 Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2006) Experience-Dependent
746 Sharpening of Visual Shape Selectivity in Inferior Temporal Cortex. *Cereb*
747 *Cortex* 16:1631–1644.

748 Gheysen F, Van Opstal F, Roggeman C, Van Waelvelde H, Fias W (2010)
749 Hippocampal contribution to early and later stages of implicit motor sequence
750 learning. *Exp Brain Res* 202:795–807.

751 Gheysen F, Van Opstal F, Roggeman C, Van Waelvelde H, Fias W (2011) The
752 Neural Basis of Implicit Perceptual Sequence Learning. *Front Hum Neurosci* 5
753 Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3213531/>
754 [Accessed March 23, 2020].

755 Handel S, Buffardi L (1969) Using Several Modalities to Perceive one Temporal
756 Pattern. *Quarterly Journal of Experimental Psychology* 21:256–266.

757 Haushofer J, Livingstone MS, Kanwisher N (2008) Multivariate Patterns in Object-
758 Selective Cortex Dissociate Perceptual and Physical Shape Similarity. *PLOS*
759 *Biology* 6:e187.

760 Hindy NC, Avery EW, Turk-Browne NB (2019) Hippocampal-neocortical interactions
761 sharpen over time for predictive actions. *Nature Communications* 10:3989.

762 Hunt RH, Aslin RN (2001) Statistical learning in a serial reaction time task: Access to
763 separable statistical cues by individual learners. *Journal of Experimental*
764 *Psychology: General* 130:658–680.

765 JASP Team (2019) JASP (Version 0.10.2)[Computer software].

766 Kaposvari P, Kumar S, Vogels R (2018) Statistical Learning Signals in Macaque
767 Inferior Temporal Cortex. *Cereb Cortex* 28:250–266.

768 Karuza EA, Emberson LL, Roser ME, Cole D, Aslin RN, Fiser J (2017) Neural
769 signatures of spatial statistical learning: Characterizing the extraction of
770 structure from complex visual scenes. *J Cogn Neurosci* 29:1963–1976.

771 Kok P, Jehee JFM, de Lange FP (2012) Less Is More: Expectation Sharpens
772 Representations in the Primary Visual Cortex. *Neuron* 75:265–270.

773 Kourtzi Z, Kanwisher N (2001) Representation of Perceived Object Shape by the
774 Human Lateral Occipital Complex. *Science* 293:1506–1509.

775 Kubovy M (1988) Should we resist the seductiveness of the
776 space:time::vision:audition analogy? *Journal of Experimental Psychology:*
777 *Human Perception and Performance* 14:318–320.

778 Lakens D (2013) Calculating and reporting effect sizes to facilitate cumulative
779 science: a practical primer for t-tests and ANOVAs. *Front Psychol* 4.

780 Lee MD, Wagenmakers E-J (2014) Bayesian Cognitive Modeling: A Practical
781 Course. Cambridge University Press.

782 Mahar D, Mackenzie B, McNicol D (1994) Modality-Specific Differences in the
783 Processing of Spatially, Temporally, and Spatiotemporally Distributed
784 Information. *Perception* 23:1369–1386.

785 Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey
786 inferotemporal cortex. *PNAS* 108:19401–19406.

787 Penney CG (1989) Modality effects and the structure of short-term verbal memory.
788 *Mem Cogn* 17:398–422.

789 Plante E, Patterson D, Gómez R, Almryde KR, White MG, Asbjørnsen AE (2015)
790 The nature of the language input affects brain activation during learning from
791 a natural language. *J Neurolinguistics* 36:17–34.

792 Repp BH, Penel A (2002) Auditory dominance in temporal processing: new evidence
793 from synchronization with simultaneous visual and auditory sequences. *J Exp*
794 *Psychol Hum Percept Perform* 28:1085–1099.

795 Richter D, de Lange FP (2019) Statistical learning attenuates visual activity only for
796 attended stimuli Frank MJ, Kahnt T, Wyart V, Heinze J, eds. *eLife* 8:e47869.

797 Richter D, Ekman M, de Lange FP (2018) Suppressed Sensory Response to
798 Predictable Object Stimuli throughout the Ventral Visual Stream. *J Neurosci*
799 38:7452–7461.

800 Saffran JR (2002) Constraints on Statistical Language Learning. *Journal of Memory*
801 and Language

802 Schapiro AC, Gregory E, Landau B, McCloskey M, Turk-Browne NB (2014) The
803 necessity of the medial temporal lobe for statistical learning. *J Cogn Neurosci*
804 26:1736–1747.

805 Schwartz O, Hsu A, Dayan P (2007) Space and time in visual context. *Nature*
806 *Reviews Neuroscience* 8:522–535.

807 Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg
808 H, Bannister PR, De Luca M, Drobniak I, Flitney DE, Niazy RK, Saunders J,
809 Vickers J, Zhang Y, De Stefano N, Brady JM, Matthews PM (2004) Advances
810 in functional and structural MR image analysis and implementation as FSL.
811 *NeuroImage* 23:S208–S219.

812 Summerfield C, Tritschuh EH, Monti JM, Mesulam M-M, Egner T (2008) Neural
813 repetition suppression reflects fulfilled perceptual expectations. *Nature*
814 *Neuroscience* 11:1004–1006.

815 Tobia MJ, Iacobella V, Davis B, Hasson U (2012a) Neural systems mediating
816 recognition of changes in statistical regularities. *NeuroImage* 63:1730–1742.

817 Tobia MJ, Iacobella V, Hasson U (2012b) Multiple sensitivity profiles to diversity and
818 transition structure in non-stationary input. *NeuroImage* 60:991–1005.

819 Tremblay P, Baroni M, Hasson U (2013) Processing of speech and non-speech
820 sounds in the supratemporal plane: Auditory input preference does not predict
821 sensitivity to statistical structure. *NeuroImage* 66:318–332.

822 Turk-Browne NB, Scholl BJ, Chun MM, Johnson MK (2009) Neural evidence of
823 statistical learning: efficient detection of visual regularities without awareness.
824 *J Cogn Neurosci* 21:1934–1945.

825 Turk-Browne NB, Scholl BJ, Johnson MK, Chun MM (2010) Implicit perceptual
826 anticipation triggered by statistical learning. *J Neurosci* 30:11177–11187.

827