Report

Current Biology

Preparatory attention incorporates contextual expectations

Highlights

- Visual cortex contains object-specific representations during search preparation
- We demonstrate this for the first time during concurrent visual scene processing
- Preparatory object representations are scaled to account for viewing distance
- Preparatory biases reflect the predicted retinal image inferred from scene context

Authors

Surya Gayet, Marius V. Peelen

Correspondence

surya.gayet@gmail.com

In brief

Attentional selection is thought to be mediated by target-specific preparatory activity in visual cortex. Gayet and Peelen provide evidence that such preparatory biases incorporate contextual expectations about object appearance, reconciling attention theories with the challenges of naturalistic vision.

Gayet & Peelen, 2022, Current Biology *32*, 1–6
February 7, 2022
https://doi.org/10.1016/j.cub.2021.11.062

CellPress

Current Biology

CellPress

Report

Preparatory attention incorporates contextual expectations

Surya Gayet[1,2,3,*] and Marius V. Peelen[1]
[1]Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6525 GD Nijmegen, the Netherlands
[2]Helmholtz Institute, Experimental Psychology, Utrecht University, 3584 CS Utrecht, the Netherlands
[3]Lead contact
*Correspondence: surya.gayet@gmail.com
https://doi.org/10.1016/j.cub.2021.11.062

SUMMARY

Humans are remarkably proficient at finding objects within complex visual scenes. According to current theories of attention,[1–3] visual processing of an object of interest is favored through the preparatory activation of object-specific representations in visual cortex.[4–15] One key problem that is inherent to real-world visual search but is not accounted for by current theories is that a given object will produce a dramatically different retinal image depending on its location, which is unknown in advance. For instance, the color of the retinal image depends on the illumination on the object, its shape depends on the viewpoint, and (most critically) its size can vary by several orders of magnitude, depending on the distance to the observer. In order to benefit search, preparatory activity thus needs to incorporate contextual expectations. In the current study, we measured fMRI blood-oxygen-level-dependent (BOLD) activity in human observers while they prepared to search for objects at different distances in indoor-scene photographs. First, we established that observers instantiated preparatory object representations: activity patterns in object-selective cortex evoked during search preparation (while no objects were presented) resembled activity patterns evoked by viewing those objects in isolation. Second, we demonstrated that these preparatory object representations were systematically modulated by expectations derived from scene context: activity patterns reflected the predicted retinal image of the object at each distance (i.e., distant search evoking smaller object representations and nearby search evoking larger object representations). These findings reconcile current theories of attentional selection with the challenges of real-world vision.

RESULTS

The vast majority of the sensory input entering through our eyes is irrelevant to our current behavioral goals. An influential idea is that behaviorally relevant visual input can be favored over irrelevant visual input by preparatory activation of neural populations that represent visual properties of the relevant (i.e., target) object, such as its color, shape, or size.[1–15] These visual properties, however, are unknown during real-world search, because they depend on the (inherently unknown) location of the target object. The retinal image size of an object, for instance, depends on the distance between observer and object. The main goal of the present study was (1) to establish whether observers generate preparatory object representations during naturalistic search where target objects produce variable retinal images and (2) to investigate whether these object representations are adjusted to match the predicted retinal image that the target object should produce at the current viewing location. Specifically, we test whether preparatory activity in visual cortex represents a small image of the target object during distant search and a large image of that same object during nearby search.

Participants were cued to search for a target object (a melon or box) in the near or far plane of an indoor-scene photograph (Figure 1A). In half of the trials, an array of objects briefly appeared at the cued distance, and observers reported whether or not the cued target object was present. Average accuracy was 67% (74% on target-present trials; 60% on target-absent trials), which was better than chance; $p < 0.0005$. All reported p values reflect the probability of incorrectly rejecting the null hypothesis (i.e., type-1 error rate), based on bootstrap tests across participants with 2,000 permutations (see STAR Methods).

The fMRI analyses reported below are based on the other half of the trials, in which no objects appeared, thus isolating activity evoked by search preparation. We focused our analyses on two main regions of interest (ROIs): object-selective cortex (OSC) (mostly corresponding to the lateral occipital complex) and early visual cortex (EVC) (mostly corresponding to V1 and V2). Previous work has shown object-specific preparatory activity in these two regions while observers prepared to search for objects in naturalistic scenes.[15] Moreover, these two regions are known to underly the processing of object size.[16–23]

Preparatory activity contains an object-specific target representation

We first needed to ascertain that object-specific representations were instantiated in visual cortex during search preparation (irrespective of search distance). In a first analysis, we tested whether search preparation for melons and boxes (following an

Please cite this article in press as: Gayet and Peelen, Preparatory attention incorporates contextual expectations, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.11.062

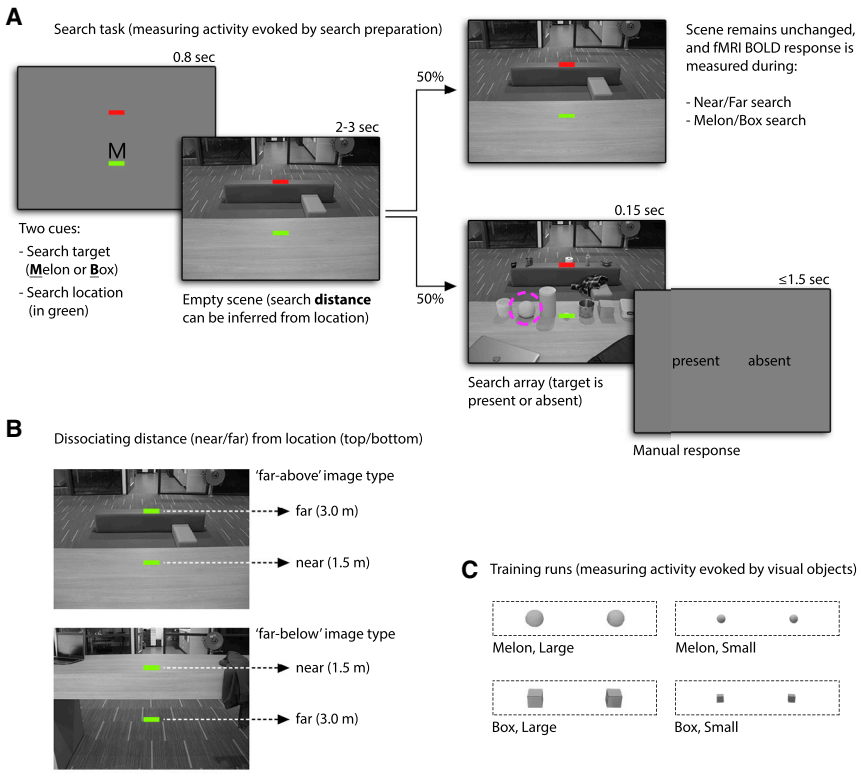

Figure 1. Methods

(A) Search task: at the start of each trial, a letter (M or B) indicated which of two possible objects participants should search for (a melon or box) and a green bar indicated the vertical location at which this target object could appear. After 800 ms, an indoor-scene photograph was presented, which—together with the location cue—enabled participants to infer the real-world distance at which the target object should appear, thus predicting the retinal size of the target object. In half of the trials, an array of objects appeared for 150 ms, following a variable delay (2 to 3 s), and participants reported (within 1.5 s) whether the target object was present or absent (here: present; circled in purple for illustrational purposes). In the other half of the trials, the scene remained unchanged for the full delay period (no objects were presented), allowing for measuring fMRI activity evoked during search preparation (for near and far melons and boxes).

(B) We used sixteen different background scenes, which categorically varied in their spatial layout, so that the "far" location could either be in the upper or lower part of the scene with equal probability. This ensured that viewing distance had to be extracted from the scene and could not be inferred from the cue alone.

(C) Training runs: isolated objects cropped from the search scenes were presented in a mini-block design. From this, we retrieved benchmark fMRI activity patterns evoked by viewing large and small melons and boxes, which could then be compared to activity patterns evoked during search preparation (for near and far melons and boxes).

"M" or "B" cue in the search task) evoked distinguishable patterns of activity, while no objects were actually presented (Figure 2A). A multivariate classification approach showed that this was the case in OSC ($p = 0.001$), but no such evidence was found for EVC ($p = 0.259$). Although in line with our hypothesis, this result could also reflect lingering activity evoked by the cues (M or B) rather than sustained visual-like representations of the target objects. To preclude contamination by cue-evoked activity, we next tested whether activity patterns evoked during search preparation (for melon versus box) could be distinguished based on benchmark activity patterns evoked by *viewing* the two target objects (i.e., images of melons versus boxes presented in isolation in separate training runs, which evoked distinguishable activity patterns in both ROIs; $p < 0.0005$ for both tests). We observed above-chance cross-classification of the target object in OSC ($p = 0.004$), confirming our hypothesis, but not in EVC ($p = 0.098$; Figure 2B). These findings replicated across a wide range of voxel inclusion thresholds (Figure 2B, rightmost panels) and after reversing the train-test direction of the cross-classification analysis (Figure S1). The generalization from visually evoked activity to preparatory activity was not driven by systematic differences in overall blood-oxygen-level-dependent (BOLD) response to melon compared to box-related conditions (Figure S2). Taken together, these findings show that activity patterns in OSC reflect a visual-like representation of the target object during search preparation. To our knowledge, this provides the first demonstration of preparatory object representations during concurrent scene processing (a requirement for real-world visual search).

Preparatory object representations are distance dependent

The main goal of this study was to find out whether preparatory object representations, as observed in OSC, are scaled to account for the current viewing distance. If this were the case, a classifier trained to distinguish between *large* melons and boxes (in the training runs) should perform best when cross-classifying target objects during *near* search (in the search task), and a classifier trained to distinguish between *small* melons and boxes should perform best when cross-classifying target objects during *far* search. In OSC, the data confirmed our hypothesis (Figure 3): the target object could be cross-classified from preparatory activity when training on size-matching objects ($p < 0.0005$), but not when training on size-mismatching objects ($p = 0.224$; difference $p = 0.017$). A similar pattern of results was observed in EVC (size-matching objects, $p = 0.040$; size-mismatching objects, $p = 0.244$), but classifier performance did not reliably differ between training regimes ($p = 0.197$). Confirming our finding in OSC, 19 out of 20 alternative voxel inclusion thresholds yielded an even more reliable difference score than the default reported above (Figure 3, top-right panel). This pattern of results was also replicated after reversing the train-test direction of the analysis

Please cite this article in press as: Gayet and Peelen, Preparatory attention incorporates contextual expectations, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.11.062

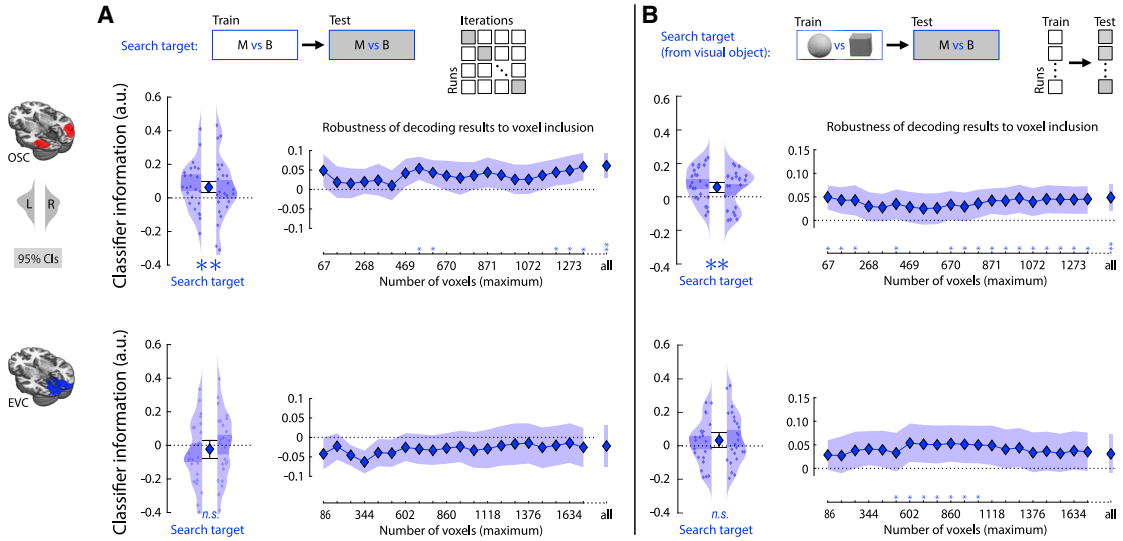

Figure 2. Object specificity of preparatory activity

(A) First, we tested whether search preparation for melons versus boxes (following an M versus B cue in the search task) evoked distinguishable patterns of activity in OSC and EVC. Multivariate classification was achieved using a linear support vector machine (libsvm)[24] on run-based beta-maps following a leave-one-run-out cross-validation procedure. To probe the robustness of the analyses conducted on the default ROIs, we repeated these analyses for twenty different voxel inclusion thresholds (rightmost plots; the default ROIs are included as "all"). Significance is reported after correction for multiple comparisons and threshold-free cluster enhancement (see STAR Methods).[25]

(B) Second, we tested whether a classifier trained on visually evoked activity in training runs (images of melons versus boxes) successfully distinguished which object (melon versus box) participants were cued to search for in the search task. Small colored dots represent classifier information (derived from distance to bound) for individual participants, obtained separately from the left and right hemispheres (displayed within the left and right kernel-density plots, respectively). The central marker reflects the population mean, averaged across hemispheres. Error bars around the central markers, shaded areas within the kernel-density plots, and shaded area in the robustness plots represent the bootstrapped 95% confidence intervals of the mean (2,000 permutations).

*p < 0.05; **p < 0.005; ***p < 0.0005. See also Figures S1–S3.

(Figure S1) but was not observed in face- and scene-selective ROIs (Figure S3). Taken together, the present findings show that preparatory activity patterns in OSC (1) resemble visually evoked representations and (2) take into account the predicted retinal image size of the target object at the current viewing distance.

DISCUSSION

The present findings demonstrate how preparatory activity can support attentional selection in the real world, where the visual input associated with an object depends on scene context: during search preparation, OSC contains a representation of the target object that flexibly adjusts to match the predicted retinal image that the target object produces at the current search location.

It is well known that expectations modulate visual processing,[26] causing expected visual objects to evoke a reduced but sharpened neural response.[27,28] Much like preparatory activity for explicit search targets, the expectation of an upcoming object suffices for evoking object-specific activity patterns in visual cortex.[29,30] In the present work, we show that observers used scene context to generate expectations about the appearance of an object. Specifically, knowing the distance at which an object was expected to appear allowed observers to generate strong predictions regarding the retinal size of the upcoming target object. Our results show that observers incorporated these predictions in their top-down attentional set, as reflected in combined distance- and object-specific preparatory activity.

We interpret the finding of distance-dependent, object-specific preparatory activity as evidence that object representations are rescaled to match the current search distance. Another interpretation of our data is the differential up or down weighting of object features based on their informativeness for the detection of target objects at a given distance. The texture of a cantaloupe melon, for instance, might be a more useful feature during near search (when texture is more visible) than during far search. Such a distance-based feature weighting mechanism could contribute to the current findings, as feature-specific preparatory activity (in the testing runs) may generalize to visually evoked activity (in the training runs): classifiers trained to distinguish large melons from large boxes may rely more strongly on texture than classifiers trained to distinguish small melons from small boxes. Crucially, both the scaling and the differential-feature weighting interpretations support the same general conclusion that object-specific preparatory activity incorporates contextual expectations about target appearance. Future research is needed to isolate the features that are most important for distance-dependent updating of preparatory activity.

The setting of the current study differs in many ways from perception in the real world.[31] The aim of the current study, however, was not to recreate real-world viewing conditions in the lab but, instead, to isolate one key aspect of real-world perception.

Please cite this article in press as: Gayet and Peelen, Preparatory attention incorporates contextual expectations, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.11.062

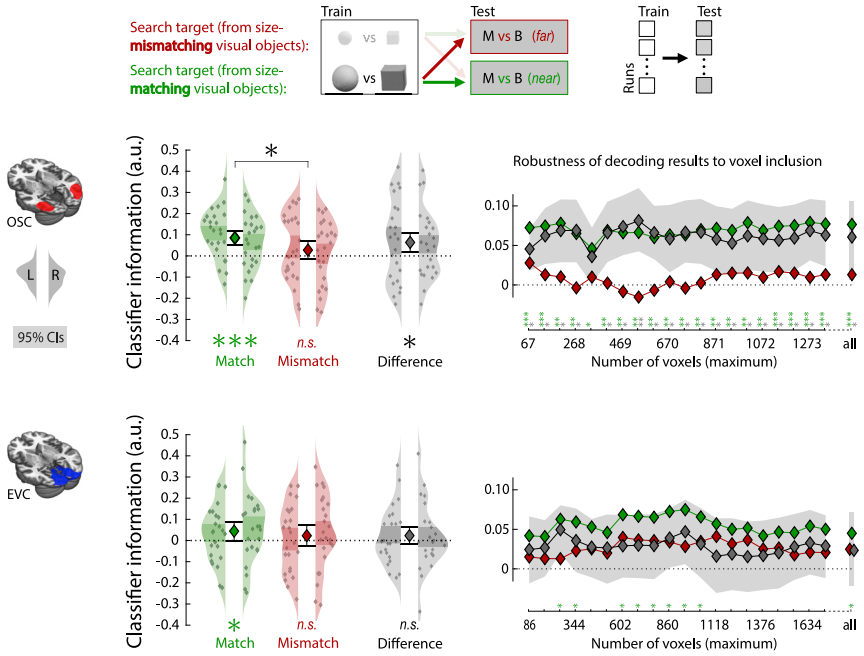

Figure 3. Size specificity of preparatory object representations

We tested whether cross-classification of target object (as shown in Figure 2B) improves when training a classifier on appropriately sized objects given the current viewing distance (e.g., large melons versus large boxes when testing on *near* search) compared to inappropriately sized objects (e.g., large melons versus large boxes when testing on *far* search). All analysis and data visualization approaches are identical to those described under Figure 2. *p < 0.05; **p < 0.005; ***p < 0.0005. See also Figures S1–S3.

In our study, viewing distance (as inferred from a photograph) determined the size with which a target object was represented during search preparation. This implies that participants were able to infer the viewing distance at which they were searching for a target object from our scene stimuli with sufficient precision, using only pictorial depth cues. Real-world vision provides observers with many more cues to obtain an even better estimate of viewing distance,[32–34] presumably making distance-dependent preparatory activity an even more effective mechanism during real-world search than in the present experiment. At the same time, it remains unknown how the present paradigm, in which observers were cued to search at two substantially distinct distances (1.5 and 3 m) within a static display, would generalize to attentional selection during dynamic visual input.

Besides the retinal image size of an object, scene context provides various other predictions that could inform object processing.[35–37] For example, the position of an object in a scene also predicts the shape of its retinal projection: a round clock on a wall to the left of the observer produces an elliptical projection on the retinae. Behavioral work has shown that, when observers searched for elliptic targets, their attention is captured by canonically circular objects (i.e., a coin) when they are rotated so as to produce an elliptic projection.[38] This shows that, at the processing stage that is relevant to visual search, objects are (at least partly) represented in terms of their retinal projection, rather than their canonical real-world shape. Consequently, predicting the shape of the retinal projection of an object based on the angle of incidence could potentially benefit search. Object-selective cortex exhibits different degrees of invariance for such location-dependent factors as illumination, viewpoint, and size.[18] Consequently, it remains unknown whether viewing distance and the angle of incidence are similarly incorporated in preparatory activity. The current work can be regarded as a proof of principle, showing that context-based predictions can be incorporated in preparatory activity, optimizing visual processing for goal-directed behavior.

There are at least two additional mechanisms that could work together with the preparatory scaling mechanism described here to support visual search at different distances. First, scene context could rapidly modulate the representations of target and distractor objects once these appear (or once they are fixated),[39] perhaps even before competition is biased by preparatory activity. Supporting this view, we recently showed that spatial attention is captured by objects whose distance-inferred size matches the size of objects currently held in short-term memory,[40] with such memory templates resembling preparatory activity.[41–43] Second, preparatory activity can be instantiated at multiple levels of the visual hierarchy, from low-level features to high-level object categories.[5] For some objects, it may thus be possible to pre-activate high-level representations that are invariant to size and viewpoint. In line with this, behavioral work showed that, when observers searched for a highly familiar object category (people or cars) in natural scenes, attention was automatically captured by outlines of a person (or car), regardless of their orientation.[44] In sum, preparatory activity can comprise multiple features, each of which might be more or less diagnostic under different circumstances, and can be instantiated at different processing levels, allowing either for relatively early image-like selection or for higher-level selection (e.g., view invariance). Considering human observers' remarkable proficiency in detecting objects in naturalistic scenes,[45–50] it is probable that we flexibly switch between search mechanisms to optimally match situational demands.

Conclusions

We proposed that, for preparatory biases to work as a selection mechanism in naturalistic conditions, these would have to (1) operate when participants concurrently process visual information (i.e., looking at a scene while preparing to search) and (2) dynamically update as a function of the current search location (to account for viewing distance, illumination, angle of incidence, etc.). Here, we show for the first time that human observers generate visual-like representations of target objects in object-selective cortex while (1) actively processing the scene that

they are searching and (2) adapting these object representations as a function of viewing distance. By doing so, we demonstrate that preparatory biases in visual cortex are a viable mechanism for visual selection in the real world.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Participants
- METHOD DETAILS
  - Apparatus
  - General experimental procedure
  - Experimental design & stimuli: Search task
  - Experimental design & stimuli: Model training
  - Experimental design & stimuli: Functional localizer
  - Acquisition of fMRI data
  - Preprocessing of fMRI data
  - Creating regions-of-interest: Object-selective cortex (OSC)
  - Creating regions-of-interest: Early visual cortex (EVC)
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Behavioral analyses
  - General linear model (GLM) estimation
  - Multivariate pattern analyses
  - Significance testing

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2021.11.062.

ACKNOWLEDGMENTS

This project received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement no. 725970, granted to M.V.P.) and from the Netherlands Organisation for Scientific Research (VI.Veni.191G.085, granted to S.G.). The authors would like to thank Nicolò Trevisan for help with the stimulus creation and data collection and Stefan van der Stigchel, Maëlle Lerebourg, Genevieve Quek, and four anonymous reviewers for helpful comments on various drafts of the manuscript.

AUTHOR CONTRIBUTIONS

S.G. and M.V.P. developed the study concept and study design. S.G. programmed the experimentation scripts and developed the pre-processing and analysis scripts in consultation with M.V.P. S.G. tested the participants. S.G. drafted the manuscript, and M.V.P. provided critical revisions. Both authors approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure that the study questionnaires were prepared in an inclusive way. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

Received: June 24, 2021
Revised: October 18, 2021
Accepted: November 25, 2021
Published: December 16, 2021

REFERENCES

1. Duncan, J., and Humphreys, G.W. (1989). Visual search and stimulus similarity. Psychol. Rev. *96*, 433–458.
2. Wolfe, J.M., Cave, K.R., and Franzel, S.L. (1989). Guided search: an alternative to the feature integration model for visual search. J. Exp. Psychol. Hum. Percept. Perform. *15*, 419–433.
3. Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual attention. Annu. Rev. Neurosci. *18*, 193–222.
4. Reynolds, J.H., and Chelazzi, L. (2004). Attentional modulation of visual processing. Annu. Rev. Neurosci. *27*, 611–647.
5. Battistoni, E., Stein, T., and Peelen, M.V. (2017). Preparatory attention in visual cortex. Ann. N Y Acad. Sci. *1396*, 92–107.
6. Beck, D.M., and Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. Vision Res. *49*, 1154–1165.
7. Chelazzi, L., Duncan, J., Miller, E.K., and Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. J. Neurophysiol. *80*, 2918–2940.
8. Chelazzi, L., Miller, E.K., Duncan, J., and Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. Nature *363*, 345–347.
9. Chawla, D., Rees, G., and Friston, K.J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. Nat. Neurosci. *2*, 671–676.
10. Giesbrecht, B., Weissman, D.H., Woldorff, M.G., and Mangun, G.R. (2006). Pre-target activity in visual cortex predicts behavioral performance on spatial and feature attention tasks. Brain Res. *1080*, 63–72.
11. Shibata, K., Yamagishi, N., Goda, N., Yoshioka, T., Yamashita, O., Sato, M.A., and Kawato, M. (2008). The effects of feature attention on prestimulus cortical activity in the human visual system. Cereb. Cortex *18*, 1664–1675.
12. Stokes, M., Thompson, R., Nobre, A.C., and Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. Proc. Natl. Acad. Sci. USA *106*, 19569–19574.
13. Puri, A.M., Wojciulik, E., and Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. Brain Res. *1301*, 89–99.
14. Esterman, M., and Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. Cereb. Cortex *20*, 1245–1253.
15. Peelen, M.V., and Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. Proc. Natl. Acad. Sci. USA *108*, 12125–12130.
16. Cohen, L., Gray, F., Meyrignac, C., Dehaene, S., and Degos, J.D. (1994). Selective deficit of visual size perception: two cases of hemimicropsia. J. Neurol. Neurosurg. Psychiatry *57*, 73–78.
17. Grill-Spector, K., Kourtzi, Z., and Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. Vision Res. *41*, 1409–1422.
18. Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., and Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. Neuron *24*, 187–203.

19. Murray, S.O., Boyaci, H., and Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. Nat. Neurosci. *9*, 429–434.
20. Eger, E., Kell, C.A., and Kleinschmidt, A. (2008). Graded size sensitivity of object-exemplar-evoked activity patterns within human LOC subregions. J. Neurophysiol. *100*, 2038–2047.
21. Konkle, T., and Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. J. Neurosci. *33*, 10235–10242.
22. Konkle, T., and Oliva, A. (2012). A real-world size organization of object responses in occipitotemporal cortex. Neuron *74*, 1114–1124.
23. Zeng, H., Fink, G.R., and Weidner, R. (2020). Visual size processing in early visual cortex follows lateral occipital cortex involvement. J. Neurosci. *40*, 4410–4417.
24. Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. *2*, 1–27.
25. Smith, S.M., and Nichols, T.E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage *44*, 83–98.
26. de Lange, F.P., Heilbron, M., and Kok, P. (2018). How do expectations shape perception? Trends Cogn. Sci. *22*, 764–779.
27. Kok, P., Brouwer, G.J., van Gerven, M.A., and de Lange, F.P. (2013). Prior expectations bias sensory representations in visual cortex. J. Neurosci. *33*, 16275–16284.
28. Kok, P., Jehee, J.F., and de Lange, F.P. (2012). Less is more: expectation sharpens representations in the primary visual cortex. Neuron *75*, 265–270.
29. Kok, P., Failing, M.F., and de Lange, F.P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. J. Cogn. Neurosci. *26*, 1546–1554.
30. Kok, P., Mostert, P., and de Lange, F.P. (2017). Prior expectations induce prestimulus sensory templates. Proc. Natl. Acad. Sci. USA *114*, 10473–10478.
31. Snow, J.C., and Culham, J.C. (2021). The treachery of images: how realism influences brain and behavior. Trends Cogn. Sci. *25*, 506–519.
32. Sprague, W.W., Cooper, E.A., Tošić, I., and Banks, M.S. (2015). Stereopsis is adaptive for the natural environment. Sci. Adv. *1*, e1400254.
33. Gibaldi, A., and Banks, M.S. (2019). Binocular eye movements are adapted to the natural environment. J. Neurosci. *39*, 2877–2888.
34. Gibaldi, A., Benson, N.C., and Banks, M.S. (2021). Crossed-uncrossed projections from primate retina are adapted to disparities of natural scenes. Proc. Natl. Acad. Sci. USA *118*, e2015651118.
35. Biederman, I., Mezzanotte, R.J., and Rabinowitz, J.C. (1982). Scene perception: detecting and judging objects undergoing relational violations. Cognit. Psychol. *14*, 143–177.
36. Oliva, A., and Torralba, A. (2006). Building the gist of a scene: the role of global image features in recognition. Prog. Brain Res. *155*, 23–36.
37. Bar, M. (2004). Visual objects in context. Nat. Rev. Neurosci. *5*, 617–629.
38. Morales, J., Bax, A., and Firestone, C. (2020). Sustained representation of perspectival shape. Proc. Natl. Acad. Sci. USA *117*, 14873–14882.
39. Sperandio, I., and Chouinard, P.A. (2015). The mechanisms of size constancy. Multisens. Res. *28*, 253–283.
40. Gayet, S., and Peelen, M.V. (2019). Scenes modulate object processing before interacting with memory templates. Psychol. Sci. *30*, 1497–1509.
41. de Fockert, J.W., Rees, G., Frith, C.D., and Lavie, N. (2001). The role of working memory in visual selective attention. Science *291*, 1803–1806.
42. Carlisle, N.B., Arita, J.T., Pardo, D., and Woodman, G.F. (2011). Attentional templates in visual working memory. J. Neurosci. *31*, 9315–9322.
43. Nobre, A.C., and Stokes, M.G. (2019). Premembering experience: a hierarchy of time-scales for proactive attention. Neuron *104*, 132–146.
44. Reeder, R.R., and Peelen, M.V. (2013). The contents of the search template for category-level search in natural scenes. J. Vis. *13*, 13.
45. Potter, M.C. (1975). Meaning in visual search. Science *187*, 965–966.
46. Potter, M.C. (1976). Short-term conceptual memory for pictures. J. Exp. Psychol. Hum. Learn. *2*, 509–522.
47. Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. J. Exp. Psychol. Hum. Percept. Perform. *7*, 604–610.
48. Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. Nature *381*, 520–522.
49. Li, F.F., VanRullen, R., Koch, C., and Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. Proc. Natl. Acad. Sci. USA *99*, 9596–9601.
50. Peelen, M.V., Fei-Fei, L., and Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. Nature *460*, 94–97.
51. Epstein, R., and Kanwisher, N. (1998). A cortical representation of the local visual environment. Nature *392*, 598–601.
52. Julian, J.B., Fedorenko, E., Webster, J., and Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. Neuroimage *60*, 2357–2364.
53. Wellcome Centre for Human Neuroimaging (2014). SPM12: Statistical parametric mapping software. https://www.fil.ion.ucl.ac.uk/spm/software/spm12/.
54. Hebart, M.N., Görgen, K., and Haynes, J.D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. Front. Neuroinform. *8*, 88.
55. Oosterhof, N.N., Connolly, A.C., and Haxby, J.V. (2016). CoSMoMVPA: multi-modal multivariate pattern analysis of neuroimaging data in Matlab/GNU Octave. Front. Neuroinform. *10*, 27.
56. Brainard, D.H. (1997). The psychophysics toolbox. Spat. Vis. *10*, 433–436.
57. Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat. Vis. *10*, 437–442.
58. Brodmann, K. (1909). Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues (Barth).
59. Wohlschläger, A.M., Specht, K., Lie, C., Mohlberg, H., Wohlschläger, A., Bente, K., Pietrzyk, U., Stöcker, T., Zilles, K., Amunts, K., and Fink, G.R. (2005). Linking retinotopic fMRI mapping and anatomical probability maps of human occipital areas V1 and V2. Neuroimage *26*, 73–82.
60. Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., and Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. Neuroimage *137*, 188–200.
61. Sassenhagen, J., and Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. Psychophysiology *56*, e13335.

Please cite this article in press as: Gayet and Peelen, Preparatory attention incorporates contextual expectations, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.11.062

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Raw data (fMRI & behavior)	This study	https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549
Processed data (fMRI & behavior)	This study	https://osf.io/nqf6p/ and https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549
Stimulus material (training and search task runs)	This paper	https://osf.io/nqf6p/ and https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549
Stimulus material (functional localizer runs)	Epstein and Kanwisher[51]	https://osf.io/nqf6p/ and https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549
Population-level regions-of-interest	Julian et al.[52]	https://osf.io/nqf6p/ and https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549
Sofftware and algorithms		
MATLAB (versions r2018a, r2021a)	Mathworks	https://www.mathworks.com/products/matlab/; RRID: SCR_001622
SPM (version 12)	Wellcome Centre for Human Neuroimaging[53]	https://www.fil.ion.ucl.ac.uk/spm/software/spm12; RRID: SCR_007037
The Decoding Toolbox, TDT (version 3.98)	Hebart et al.[54]	https://sites.google.com/site/tdtdecodingtoolbox/ ; RRID: SCR_017424
CosmoMVPA (last git pull: 16-12-2020)	Oosterhof et al.[55]	http://cosmomvpa.org; RRID: SCR_014519
Psychtoolbox (version 3.0.14)	Brainard[56] and Pelli[57]	http://psychtoolbox.org/; RRID: SCR_002881
Image Processing Toolbox (version r2018a)	Mathworks	https://nl.mathworks.com/products/image.html
Other		
MATLAB code for running experiment	This study	https://osf.io/nqf6p/ and https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549
MATLAB code for fMRI data pre-processing, region-of-interest creation, analyses, and statistical testing	This study	https://osf.io/nqf6p/ and https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549
MATLAB code for analysis of behavioral data	This study	https://osf.io/nqf6p/ and https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Surya Gayet (surya.gayet@gmail.com or s.gayet@uu.nl). This study did not generate new unique reagents.

Materials availability

All stimuli are publicly available at the Open Science Framework (OSF) project page associated with this study (https://osf.io/nqf6p/) and the Donders Institution repository project page (https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549), which are listed in the Key resources table. Both project pages are publicly available per the date of publication, but note that access to the Donders Institution repository requires a one-time registration.

Data and code availability

- Behavioral data and processed fMRI data (beta-maps and regions of interest in MNI space) have been deposited at the OSF project (https://osf.io/nqf6p/) and the Donders Institution repository (https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549) listed in the Key resources table, both of which are publicly available. Raw fMRI data in native subject space (activity maps and defaced structural scans) and pre-processed fMRI data (registered activity maps normalized to MNI space) are

deposited only at the Donders Institution repository. Privacy regulations require one-time registration to access these data in the Donders Institution repository.

- Experiment scripts and analysis files (accompanied by comprehensive read-me files, and in-script annotations) have been deposited at the OSF project (https://osf.io/nqf6p/) and the Donders Institution repository (https://data.donders.ru.nl/collections/di/dcc/DSC_2018.00046_549) listed in the Key resources table, both of which are publicly available.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Participants

Participants were recruited through the Radboud university participant pool (SONA systems) and participated for monetary reward, after providing informed consent. The study was in accordance with the institutional guidelines of the local ethical committee (CMO region Arnhem-Nijmegen, the Netherlands, Protocol CMO2014/288).

A total of 24 participants (12 females, mean age = 24.1, SD = 5.2) took part in this study, and none were excluded. All participants completed two two-hour experimental sessions on separate days. The predetermined samplesize of 24 followed from a trade-off between (A) the sample size to achieve 80% power for obtaining an effect of medium size (N = 34), and (B) our preference for having more within-subject power (i.e., two sessions instead of one), totaling to 48 experimental sessions. Maximizing the within-subject power was deemed necessary, because the hypothesized object-selective responses in preparatory activity were expected to yield small effect sizes, and key (decoding) analyses were performed within participants.

METHOD DETAILS

Apparatus

Participants viewed the stimuli through a mirror mounted on the head coil of the scanner. Stimuli were presented on a 1024 × 768 EIKI LC – XL100 projector (60 Hz refresh rate), back-projected onto a projection screen (Macada DAP diffuse KBA) attached to the back of the scanner bore. The effective viewing distance (eyes-mirror + mirror-screen) approximated 1440mm. Participants provided responses with the index fingers of the left and right hand, on a HHSC-2x4-C button box in each hand, connected to the serial port of the computer handling the stimulus presentation. All stimulus materials and experimental scripts described below can be found in the online repository, listed in the Key resources table.

General experimental procedure

Upon arrival at the scanner facilities, participants were guided through the three different tasks that they would perform inside the scanner (i.e., the search task, the training runs, and the functional localizer runs), as described below. Instructions were provided verbally, accompanied by a click-through demo and followed by a brief practice session on the experimenter's laptop, to verify that participants understood the task instructions before going into the scanner. In the scanner, participants practiced the search task during the five-minute anatomical scan, while the experimenter monitored the behavioral responses. Participants performed a total of 24 functional runs. Each functional run started and ended with 15 s of fixation.

Experimental design & stimuli: Search task

The search task was designed to measure fMRI BOLD responses evoked during search preparation for different types of objects. Participants were instructed that they would search for one of two specific objects in scene photographs; a Cantaloupe melon and a small cube-shaped cardboard box (which was physically present during the instructions). On each trial, the letter 'M' or 'B' indicated whether they had to search for a melon or box, respectively. A green marker would indicate at what vertical location of the upcoming scene the target object could appear. After 1600ms, the scene appeared and, after a variable delay, an array of objects was briefly (150ms) presented within the scene. The array consisted of 14 to 20 objects, of which 12 to 18 distractor objects (drawn from of a pool of about 35 objects) such as a sweater, a laptop, a tea mug, a plant, and a watering can. The entire scene was then removed and backward masked (100ms), and participants were required to swiftly respond whether the target object was present or absent (within 1500ms), using button boxes in the left and right hand. After the response they would get feedback (correct, incorrect, or too slow), and the fixation dot changed color to announce the start of the next trial (1500ms inter-trial-interval). Participants were instructed that on half of the trials, no array of objects would appear within the scene, and hence no response would be required. Finally, it was stressed that participants were required to maintain fixation on the green marker throughout the trial, to ensure that they could detect the target object irrespective of its horizontal location (i.e., left or right of fixation). The two possible target locations always contained an object: either the target object, or a foil. The foil was either the non-target object (e.g., the melon when the box was cued) or an object that matched the shape (but not the size) of the target object (e.g., a football when the melon was cued on the far plane).

The study comprised sixteen search task runs, totaling to 512 trials per participant. Each search task run consisted of 32 trials, in an event-related design. Within each of these runs, five factors were counterbalanced within-participant and presented in random order. First, a green marker cued participants to fixate the upper part or the lower part of the upcoming scene (see Figure 1A). Second, one

Please cite this article in press as: Gayet and Peelen, Preparatory attention incorporates contextual expectations, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.11.062

of two distinct types of scenes could be presented: scenes in which the far region was in the upper part of the image or the lower part of the image (see Figure 1C). Jointly, these two factors determined the search distance (i.e., near or far). Third, one of two possible object cues could be presented (i.e., the letter 'M' or 'B' indicating whether participants had to search for the presence of a melon or box respectively. Fourth, after a variable delay an array of objects appeared, or not. Fifth, on those trials in which an array of objects appeared, the cued target object could either be present in the array or not. One additional factor was counterbalanced with the five factors listed above across all eight search task runs of an experimental session, but not within each run: the scene background was chosen from one of eight scenes per scene type (sixteen scenes in total). Then a number of factors were (maximally) equated within runs, but could not be counterbalanced with the factors listed above. First, the type of distractor object at the non-target location could either be of a shape or a size that was similar to the cued target object. Second, the variable asynchrony between the scene onset and the onset of the search array could either be 2, 2.5 or 3 s (on trials in which no array was presented, the trial ended after the longest delay had elapsed). Finally, the mapping of the present/absent responses with the left/right button boxes was counterbalanced across participants. Participants never saw the same array of objects more than once.

The indoor scenes were photographed at sixteen different indoor locations within the Radboud University campus (Nijmegen, the Netherlands), using a digital camera on a tripod. A custom-made script was used to compute object positioning and camera angle, ensuring that the target objects would produce the same two retinal image sizes (large and small, corresponding to viewing distances of 1.5 and 3 m), at the same eccentricity (left and right of fixation, in the near and far regions of the scene) in all sixteen different scene families. By doing so we ensured that the (predicted) sizes of the different target objects would be nearly identical within each of the two distance conditions (thus allowing for training binary classification algorithms). The images were turned grayscale, and any text appearing in the photographs was blurred using Photoshop 2017 (Adobe Inc.). The masks consisted of Pink (1/f filtered) noise, generated prior to each trial. The scene stimuli subtended 20 (width) by 14.4 degrees of visual angle (dva), and the target objects subtended about 0.9 by 0.9 dva (far) and 1.8 by 1.8 dva (near), with their inner edges positioned at an eccentricity of 3.2 dva from fixation. The upper and lower fixation positions were separated by 5.4 dva.

All sixteen different scenes (without objects), and all 512 unique search arrays used in the present experiment are publicly available via the Open Science Framework project page listed in the Key resources table.

Experimental design & stimuli: Model training

The purpose of the training runs was twofold: (1) retrieving benchmark activity patterns evoked by viewing isolated objects to train the multivariate models, and (2) identifying voxels that are visually responsive to the specific target stimuli used in our study (i.e., for defining the early visual cortex ROIs).

Participants were instructed to fixate the center of the screen, while pairs of objects were simultaneously presented to the left and right of fixation. Their task was to press a button whenever one of the two objects was 20% smaller or 20% larger than the remainder of the stimuli (i.e., an oddball detection task). The presentation of objects was subdivided into series (mini-blocks) comprising specific image categories (namely large and small melons and boxes), but this was irrelevant to the participants' task.

Training runs consisted of a mini-block design, and the study comprised four training runs. Each run comprised 16 mini-blocks (four repetitions of four conditions, followed by a baseline fixation block). Each mini-block lasted 14.7 s and comprised 20 unique images (presented for 450 ms each, with a 250 ms blank in between). These images consisted of the target objects from the search tasks (i.e., melons and boxes) that were cropped from the scenes, equated in luminance, and presented on a uniform gray background. The objects were presented in pairs, one to the left and one to the right of fixation, at the exact same position (relative to fixation) as the target objects in the search task runs.

Experimental design & stimuli: Functional localizer

The purpose of the functional localizer runs was to identify object-selective voxels in individual participants. Participants were instructed to fixate the center of the screen, while different images were presented at fixation. Their task was to press a button whenever any image was presented twice in succession (i.e., a 1-back task). The presentation of images was subdivided into series (mini-blocks) comprising specific image categories (namely objects, scrambled objects, faces, and houses / landscapes), but this was irrelevant to the participants' task.

The design of the localizer runs was identical to the design of the training runs, and the study comprised a total of four localizer runs. The stimuli and design of the localizer runs were based on (and nearly identical to) that of Epstein and Kanwisher.[51] In our set-up, the stimuli subtended 12 by 12 dva.

Acquisition of fMRI data

fMRI data were acquired on a 3T Magnetom PrismaFit MR Scanner (Siemens AG, Healthcare Sector, Erlangen, Germany) using a 32-channel head coil. A T2*-weighted gradient echo EPI sequence with 6x multiband acceleration factor was used for acquisition of functional data (TR 1 s, TE 34ms, flip angle 60°, 2 mm isotropic voxels, 66 slices). For the search task, 295 images were acquired per run and 318 images were acquired per run for the training and localizer runs. A high-resolution T1-weighted anatomical scan was acquired at the start of each experimental session, using an MPRAGE sequence (TR 2.3 s, TE 3.03ms, flip angle: 8°, 1 mm isotropic voxels, 192 sagittal slices, FOV 256 mm).

Please cite this article in press as: Gayet and Peelen, Preparatory attention incorporates contextual expectations, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.11.062

Preprocessing of fMRI data

Data preprocessing was performed using SPM12. Preprocessing steps included field-map correction, two-step spatial realignment of the functional images, normalization to MNI 152 space (no down-sampling), and smoothing with a 3mm (FWHM) Gaussian filter. The two experimental sessions were independently warped into MNI space, and then combined.

Creating regions-of-interest: Object-selective cortex (OSC)

We ran a general linear model to model the responses evoked by viewing intact objects and scrambled objects in the localizer runs. Individual mini-blocks were modeled as boxcars and convolved with the canonical hemodynamic response function (HRF) provided in SPM12, and six motion parameters were included as nuisance regressors. Next, we computed a univariate contrast on the resulting run-based beta-maps to identify voxels that exhibited a significantly ($p_{uncorrected} < 0.05$) stronger response to intact objects than to scrambled objects. The ensuing collection of object-selective voxels for each participant was then intersected with a population-level functionally-defined object-selective mask (retrieved from Julian et al.[52]), constraining the ROIs to lateral occipital regions, including the lateral occipital complex. The size of the OSC ROI varied across participants, with an average size of 1497 voxels (SD = 672; range = [181, 2693]) in the left hemisphere, and of 1173 voxels (SD = 512; range = [148, 2377]) in the right hemisphere.

To assess the robustness of our results, we created twenty additional ROIs for each participant, in which we differently constrained the maximum number of object-selective voxels. To this end, we sorted participants' object-selective voxels within the group-level mask from most to least object-selective. Then, we created ROIs by keeping only the N most object-selective voxels within the mask, where "N" was a number increasing from zero to the median number of (significantly) object-selective voxels across all participants in twenty equidistant steps. Thus, this resulted in twenty additional OSC sub-ROIs of increasing size, and with increasingly liberal voxel inclusion.

All ROIs were initially constrained to a single hemisphere, allowing us to perform classification analyses within the left and right hemisphere separately (and averaging the results across hemispheres later).

Creating regions-of-interest: Early visual cortex (EVC)

We ran a general linear model to model the responses evoked by viewing target objects in the training runs (large and small melons and boxes). Individual mini-blocks were modeled as boxcars and convolved with the canonical HRF, and six motion parameters were included as nuisance regressors. Next, we computed a one-sample univariate contrast on the resulting run-based beta-maps to identify voxels that exhibited a significantly ($p_{uncorrected} < 0.05$) stronger response to all four object categories in the training runs (i.e., large and small melons and boxes) relative to the implicit baseline. The ensuing collection of visually responsive voxels for each participant was then intersected with an anatomical mask constituted of Brodmann's Areas 17 and 18[58], thus constraining the ROIs to a brain region mostly corresponding to primary and secondary visual cortex[59]. The EVC ROI had an average size of 1713 voxels (SD = 604; range = [540, 2968]) in the left hemisphere, and of 1939 voxels (SD = 747, range = [690, 3244]) in the right hemisphere. Akin to the approach described above for the OSC ROI, we created twenty additional EVC sub-ROIs per participant that included an incremental number of visually responsive voxels. Again, separate ROIs were created for the left and right hemisphere.

QUANTIFICATION AND STATISTICAL ANALYSIS

Behavioral analyses

The search task comprised a total of 256 trials that required a behavioral response (i.e., in which an array of objects was presented), half of which were target-present trials, and half of which were target-absent trials. Trials that did require a response but in which no response was provided within the time limit were excluded from further behavioral analysis, so that the eventual analysis comprised an average of 126 (SD = 3.3) target-present trials, and 125.8 (SD = 3.4) target-absent trials per participant. Participants were 73.7% (SD = 9.7%) accurate on the remaining target-present trials (corresponding to a hit rate of 0.74), and 60.1% (SD = 13.3) accurate on target-absent trials (corresponding to a false-alarm rate of 0.40). To test for above-chance performance at the group level, we subtracted individual participants' false alarm rates from their hit rates, to obtain a single performance metric for each participant. Then, on each of 2,000 iterations, we drew N values from the resulting performance metrics, and computed the arithmetic mean (with N corresponding to the sample size of 24). To test for significant above chance-level performance, we computed the fraction of iterations (out of 2,000) that yielded a positive average performance at the group level. If a positive value was observed on more than 95% of iterations, this was regarded as significant above chance performance, given an alpha level of 0.05.

Using the same general analysis approach, we observed no difference in behavioral performance between trials in which the target was a melon and trials in which the target was a box, $p = 0.873$. There was a small difference in behavioral performance between near targets (69% correct) compared to far targets (65% correct), $p = 0.048$, and between scenes in which the far plane was above the near plane (69% correct) compared to scenes in which the far plane was below the near plane (65% correct), $p = 0.0016$.

In the training runs, participants correctly reported 79% of size changes (i.e., hit rate) within our predefined response deadline of 1.5 s. Behavioral performance was slightly lower for melons (78% correct) than for boxes (80% correct), $p = 0.0139$, but did not differ between small and large objects, $p = 0.810$. It should be noted that no performance difference (e.g., between near-far or melon-box conditions) was shared between the search task runs and the training runs. It is therefore unlikely that successful cross-classification capitalized on differences in task difficulty.

Please cite this article in press as: Gayet and Peelen, Preparatory attention incorporates contextual expectations, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.11.062

General linear model (GLM) estimation

To model responses evoked by viewing target objects in the model training runs (large and small images of melons and boxes), we ran a general linear model on the data of each participant. Individual mini-blocks were modeled as boxcars and convolved with the canonical hemodynamic response function provided in SPM12. The GLM captured four conditions of interest, based on the factors 'object size' and 'object type': large images of melons, large images of boxes, small images of melons, and small images of boxes. To model responses evoked during search preparation in the search task, the search delays of individual trials (from scene onset to scene offset) were modeled as boxcars and convolved with the canonical HRF. Note that this included only the 50% of trials in which no array of objects appeared, and the scene thus remained unchanged and devoid of objects. For each participant, a single GLM was used to model all sixteen runs across two scanning sessions. This GLM captured the four conditions of interest based on the factors 'search distance' and 'object cue': near search for melons, near search for boxes, far search for melons, far search for boxes. In all GLMs, six motion parameters and one run-based regressor were included as nuisance regressors, and betas were estimated on a run-basis.

Multivariate pattern analyses

All multivariate classification analyses were performed with The Decoding Toolbox[54] using a linear support vector machine (hereafter SVM; libsvm[24]). Classification analyses were performed on the run-based beta weights obtained from the GLMs (described in the previous paragraph), and were conducted within the left and right hemispheres separately. The results from both hemispheres were combined in the final step just prior to statistical testing. A leave-one-out cross-validation approach was used for the within run-type classification (i.e., within search task, or within model training runs), and a single-step cross-classification approach was used for the cross-classification analyses (e.g., training on visually evoked activity from the model training runs, and testing on activity evoked during search preparation in the search task runs).

For classification of object cue (melon versus box in the search task; Figure 2A), object type (melon versus box in the training runs), and cross-classification of object cue from object type (Figure 2B), all melon versus box maps were used, ignoring whether they were large or small, far or near. For the main analyses depicted in Figure 3 (cross-classification of object cue from size-matching and size-mismatching objects in the training runs) two separate classifiers were trained. First, a classifier was trained to distinguish between *small* melons and boxes, and tested on search preparation for melons versus boxes in the *far* plane (matching condition) and in the *near* plane (mismatching condition). Second, a classifier was trained to distinguish between large melons versus large boxes, and tested on search preparation for melons versus boxes on the *near* plane (matching condition) and in the *far* plane (mismatching conditions). The two matching conditions were then combined, and the two mismatching conditions were combined.

The number of training examples per classification analysis depended on the type of run used to train the classifier (there were 16 search task runs, and 4 model training runs), and the number of conditions included in the classification analysis (4 in total: large/near melon, small/far melon, large/near box, small/far box). To illustrate, we used 16 training examples per classification analysis to distinguish between visual presentation of melons versus boxes in the model training runs: 8 melon beta-maps and 8 box beta-maps (collapsed over size). Similarly, there were 64 test examples for the main classification analyses: 32 melon beta-maps and 32 box beta-maps.

To quantify the amount of information that is present in patterns of neural activity as retrieved by SVM classification (e.g., information about object type in OSC) we derived the following metric of classifier information from the distance-to-bound values *D*:

$$Classifier\ information = \frac{rank\_number\left(\overrightarrow{D}\right)\cdot\overrightarrow{L}}{(1:N)\cdot\overrightarrow{L}} \quad \text{(Equation 1)}$$

where N refers to the number of test examples for a given classification analysis (e.g., 64 for melon versus box classification in search task runs); D is an array of length N comprising the distance-to-bound values for each of the N test examples; and L is an array of length N containing the N labels (i.e., either −1 or 1) describing the correct classification category of the corresponding values in D (e.g., "-1" for "melon" and "1" for "box"). Put simply, the numerator of this equation reflects the amount of classifier information (positive for correct classification and negative for incorrect classification), and the denominator is a constant, which normalizes the classifier information metric to the range [-1, 1].

The sign of the distance-to-bound values reflects binary classification decision, with negative and positive values reflecting classifier decision "melon" and "box," respectively. The magnitude of these values can be regarded as the classifier's certainty of this classification decision. By computing the dot product of these distance-to-bound values (negative for "melon" and positive for "box"), with their associated correct test labels ("-1" for "melon" or "1" for "box"), the ensuing metric provides an increasingly positive value for increasingly correct classification and an increasingly negative value for increasingly incorrect classification. Instead of using the raw distance-to-bound values, we use their rank values (where the most negative value is ranked 1, and the most positive value is ranked N). This is important, because the hyperspaces obtained from different linear classifiers are not directly comparable (e.g., when averaging across hemispheres or conditions of non-interest, or comparing across conditions of interest and participants). Using this approach, we remain agnostic to the shape or space of the different hyperspaces, while still penalizing incorrect classifications more when they are very 'certain' (e.g., rank 1 or 64 out of 64) more than when they are very uncertain (e.g., rank 32 or 33 out of 64).

Please cite this article in press as: Gayet and Peelen, Preparatory attention incorporates contextual expectations, Current Biology (2021), https://doi.org/10.1016/j.cub.2021.11.062

This metric of classifier information has two crucial advantages over traditional binary classification accuracy. First, continuous (i.e., distance-to-bound) measures are more sensitive than binary classification decisions[60]. Second, this approach is inherently immune to classification biases that systematically favor one label over the other. Importantly, because our metric relies on rank order, it does share the key advantage of binary classification metrics: being agnostic as to the shape or range of the hyperplane of the linear classifier, thus allowing for comparison across multiple classifiers.

Significance testing

For each main ROI (e.g., OSC), and each condition of interest (e.g., classification of object type within-training runs) we performed bootstrap tests against chance, with 2,000 permutations (allowing for a lower bound of $p_{MIN} < 0.0005$). Specifically, on each iteration, we drew 24 samples with replacement from the 24 classifier information metrics (i.e., one for each participant), and computed the arithmetic mean. To test for significance, we then computed the fraction of iterations (out of 2,000) that yielded above chance-level classifier information at the group level. If a positive value was observed on more than 95% of iterations, this was regarded as significant above chance-level classifier performance, given an alpha level of 0.05. Note that this approach entails a directional test, which followed from the strong prediction that classifier information should be either at chance or above chance but not below chance.

To evaluate the robustness of our results observed in our primary ROIs, we conducted each classification analysis in each generic ROI (EVC and OSC) an additional twenty times, in the twenty sub-ROIs of increasing size (see ROI description above). In all twenty additional analysis, a new SVM was trained and tested (following the same classification procedure described above) and classifier information metrics were computed as described above. In addition, we also computed classifier information metrics after pseudo-randomly permuting the correct test-labels (e.g., melon or box) against which the classifier outcomes were pitted. Importantly, the same (permuted) labels were used across all twenty sub-ROIs, to preserve inherent correlations between sub-ROIs. This procedure was repeated 2,000 times to generate a null distribution that has the same variance and autocorrelations as the actual data, but should not carry any information about object type. Next, we applied threshold-free cluster enhancement[25] (TFCE) using the CoSMoMVPA toolbox[55]. TFCE boosts belief in consecutive data-points with signal (i.e., representing cluster-like spatial support of individual data-points). TFCE-scores were computed for the observed data as well as the null data. The eventual test statistic, as reported in the figures, conveys how likely a given TFCE-score is (for each sub-ROI), given the maximum TFCE-score across all sub-ROIs in the null data, thus accounting for cumulative Type I error. Importantly, using this method, statistical significance establishes the existence of above-chance classification; it does not allow for making claims about the location or extent (i.e., sub-ROIs) of this effect[61].