




The occipital place area represents visual information about walking, not crawling

Christopher M. Jones , Joshua Byland , Daniel D. Dilks *

Department of Psychology, Emory University, Atlanta, GA 30322, United States

* Corresponding author: Department of Psychology, Emory University, Atlanta, GA 30322, United States. Email: dilks@emory.edu

Recent work has shown that the occipital place area (OPA)—a scene-selective region in adult humans—supports “visually guided navigation” (i.e. moving about the local visual environment and avoiding boundaries/obstacles). But what is the precise role of OPA in visually guided navigation? Considering humans move about their local environments beginning with crawling followed by walking, 1 possibility is that OPA is involved in both modes of locomotion. Another possibility is that OPA is specialized for walking only, since walking and crawling are different kinds of locomotion. To test these possibilities, we measured the responses in OPA to first-person perspective videos from both “walking” and “crawling” perspectives as well as for 2 conditions by which humans do not navigate (“flying” and “scrambled”). We found that OPA responded more to walking videos than to any of the others, including crawling, and did not respond more to crawling videos than to flying or scrambled ones. These results (i) reveal that OPA represents visual information only from a walking (not crawling) perspective, (ii) suggest crawling is processed by a different neural system, and (iii) raise questions for how OPA develops; namely, OPA may have never supported crawling, which is consistent with the hypothesis that OPA undergoes protracted development.

Key words: visually guided navigation; locomotion; scene processing; OPA; inferior parietal lobule (IPL).

Significance statement

Our ability to navigate through the local visual environment (e.g. moving through a kitchen without running into the kitchen walls or banging into the kitchen table)—a process we refer to as “visually guided navigation”—is the foundation of many of our essential everyday behaviors. Here, we show that a scene-selective cortical region (i.e. the occipital place area [OPA]), known to be involved in visually guided navigation, intriguingly represents visual information from only 1 perspective by which humans move through their local visual environments (i.e. walking) and not from the perspective by which we had done so much earlier in life (i.e. crawling)—providing a deeper understanding of the systems underlying our critical ability to navigate our world.

Introduction

Moving about the local visual environment, avoiding boundaries and obstacles—a process we refer to as “visually guided navigation”—is a fundamental component of daily life and the bedrock of virtually all independent behaviors. Perhaps, not surprising then, it has been hypothesized that a scene-selective cortical region in adult humans—the OPA (Dilks et al. 2013)—is specifically involved in visually guided navigation (Dilks et al. 2022). Indeed, several fMRI studies found that the OPA (and not the 2 other scene-selective regions, the parahippocampal place area [PPA] and the retrosplenial complex [RSC]) represents at least 4 kinds of information relevant for visually guided navigation: (i) “sense” (left/right) information (Dilks et al. 2011); (ii) egocentric distance (near/far) information (Persichetti and Dilks 2016); (iii) local scene elements (“parts”), including boundaries (e.g. walls

and obstacles (e.g. furniture; Kamps, Julian, et al. 2016; Dillon et al. 2018; Henriksson et al. 2019; Park and Park 2020; Cheng et al. 2021); and (iv) possible routes through a local scene (Bonner and Epstein 2017; Persichetti and Dilks 2018).

Perhaps, even more comprehensive though is another fMRI study (Kamps, Lall, et al. 2016) investigating the response in OPA to videos mimicking the actual first-person visual experience of walking through a local environment, encompassing all of the above navigationally relevant information plus first-person perspective motion. Consistent with the hypothesized role of OPA in visually guided navigation, this study found that the OPA responded more to the videos mimicking walking through a local environment than to static images taken from the very same movies, rearranged such that the walking perspective was disrupted. However, humans actively move about their immediately visible environments well before they walk, actually beginning with crawling as an infant.

Thus, here, we ask whether OPA represents visual information from the 2 perspectives by which humans move about their local environment (i.e. crawling followed by walking), or instead—and perhaps counterintuitively—represents visual information about walking only, and not crawling. Perhaps, this latter possibility, is actually not so counterintuitive when considering that walking and crawling are actually fundamentally different types of locomotion. For example, when we walk, we use only our legs, but when we crawl, we use both our legs and our arms (and arguably, arms may be more important for human crawling than legs; Adolph et al. 1998). Considering both of these possibilities then, what is the precise role of OPA in visually guided navigation?

To directly address this question, we compared the response in OPA to videos depicting the actual first-person visual experience

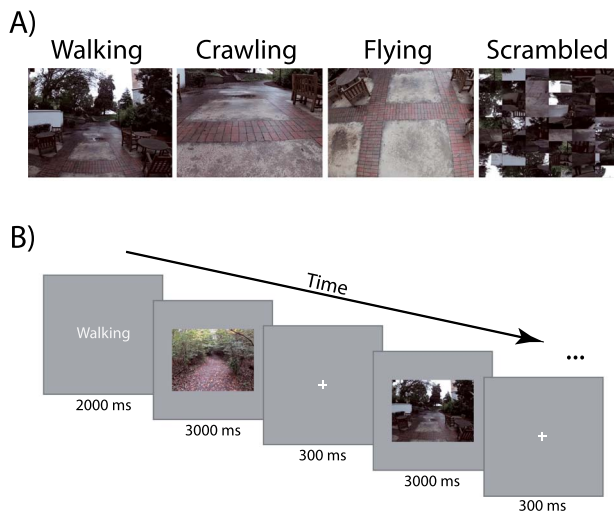


Fig. 1. A) Example frames from the “walking,” “crawling,” “flying,” and “scrambled” videos. B) Visualization of stimulus presentation procedure. Participants were first shown a word for 2 s indicating the perspective of the upcoming videos (“walking,” “crawling,” “flying,” and “scrambled”). Then, each video was played for 3 s followed by a 300-ms interstimulus interval between each video, resulting in a total block length of 19.8 s.

from the 2 perspectives by which humans actively move about the local environment (i.e. “walking” and “crawling”) versus 2 other videos by which humans do not (i.e. a top-down arial perspective or “flying”) or cannot move about the local environment (i.e. “scrambled” versions of the walking videos) (Fig. 1A). If OPA represents visual information from the 2 perspectives by which humans move about their environment, then it will respond significantly more to both the walking and crawling videos compared to the flying and scrambled ones. By contrast, if OPA represents visual information from a walking perspective only, then it will respond significantly more to the walking videos than to any of the other videos, including crawling.

Materials and methods

Participants

Fifteen participants (ages = 21–29, mean age = 24.6; 9 females, 6 males) were recruited for this experiment. All participants had normal or corrected-to-normal vision and reported no history of neurological conditions. All participants gave informed consent and were compensated for their participation.

Experimental design

For our primary analyses, we used a region of interest (ROI) approach in which we first localized scene-selective ROIs (Localizer Runs) and then used an independent set of runs to investigate the responses in each ROI to videos depicting the actual first-person visual experience of moving through local environments—from either a “walking” or “crawling” perspective as well as 2 control conditions: “flying” and “scrambled” (Experimental Runs). For both the Localizer and Experimental Runs, participants performed a 1-back task, responding every time the same image was presented twice in a row.

For the Localizer Runs, ROIs were identified using a standard method described previously (Epstein and Kanwisher 1998). Specifically, a blocked design was used in which participants viewed static images of faces, scenes, objects, and scrambled

objects. Each participant completed 2 Localizer Runs. Each run was 336 s long and consisted of 4 blocks per stimulus category. For each run, the order of the first 8 blocks was pseudorandomized, and the order of the remaining 8 blocks was the palindrome of the first 8. Each block contained 20 images from the same category for a total of 16-s blocks. Each image was presented for 300 ms, followed by a 500 ms interstimulus interval, and subtended 8° by 8° of visual angle. We also included 5 16-s fixation blocks: 1 at the beginning, 3 in the middle interleaved between each palindrome, and 1 at the end of each run.

For the Experimental Runs, we made a total of 12 3-s video clips for each of our experimental conditions (“walking,” “crawling,” “flying,” and “scrambled”). These videos were filmed using a GoPro camera. For the walking videos, the videos were taken while 1 of the authors (JB)—with the camera attached to his forehead—walked through 12 different places (e.g. a backyard, a parking lot, and a hallway). For the crawling videos, the videos were taken while JB—again, with the camera attached to his forehead—crawled through the same 12 places in which the walking videos were filmed. For the flying videos, the GoPro camera was mounted on a rod and held approximately at 10 feet in the air, facing down at the ground rather than facing out as if walking, while JB walked through the same 12 places again. (Note that our goal was not to simulate flying but to rather show a video perspective by which humans do not navigate, while keeping the same “scene” information in the flying videos compared to the walking and crawling ones.) Finally, for our scrambled videos, we divided our walking videos into a 9×9 grid and randomly shuffled the cells within the grid to scramble the video. The scrambled order of the 9×9 cells remained the same throughout each video clip, and the temporal order was kept intact. All the video clips subtended approximately $15.7^\circ \times 20.7^\circ$ visual angle. All participants completed 8 Experimental Runs, however, for 2 participants, only 7 Experimental Runs were included in analysis (due to the videos freezing in the first run of one participant, and data corruption in the first run in another participant). Each run was 450 s long and consisted of 4 blocks of 6 videos for each perspective condition. For each run, the order of the first 8 blocks was pseudorandomized, and the order of the remaining 8 blocks was the palindrome of the first 8. Before each block, a word was presented for 2 s, indicating the perspective of the upcoming videos (either “walking,” “crawling,” “flying,” or “scrambled”). Participants were instructed to imagine themselves navigating through each environment from the indicated perspective. Then, each video was played for 3 s followed by a 300-ms interstimulus interval between each video, resulting in a total block length of 19.8 s (Fig. 1B). We also included 5 19.8-s fixation blocks: 1 at the beginning, 3 in the middle interleaved between each palindrome, and 1 at the end of each run.

fMRI scanning

All scanning was performed on a 3T Siemens Trio scanner in the Facility for Education and Research in Neuroscience at Emory University. Functional images were acquired using a 32-channel head matrix coil and a gradient-echo single shot echoplanar imaging sequence (32 slices, TR = 2S, TE = 30 ms, voxel size = $3 \times 3 \times 3.6$ mm, and a .25 interslice gap). For all scans, slices were oriented approximately between perpendicular and parallel to the calcarine sulcus, covering all of the occipital and parietal lobes, as well as most of the temporal lobe. Whole-brain, high-resolution anatomical images were also acquired for each participant for use in registration and anatomical localization.

Data analysis

Analysis of fMRI data was conducted using the FSL software (Smith et al. 2004) and the Freesurfer Functional Analysis Stream FS-FAST; (<http://surfer.nmr.mgh.harvard.edu/>). ROI analyses were done using the FS-FAST ROI toolbox. Before statistical analysis, images were motion-corrected (Cox and Jesmanowicz 1999), detrended, and fit using a double gamma function. fMRI data processing was done in 2 stages in FSL: first level, looking at each run per person, and at second level, combining the runs in that person. Moreover, the motion parameters from motion correction were included in the first-level regression model as nuisance variables. Next, localizer data, but not experimental data, were spatially smoothed with a 5-mm kernel. Scene-selective regions, OPA, PPA, and RSC, were bilaterally defined in each participant (using data from the independent Localizer scans) as those regions that responded more strongly to scenes than objects ($P < 10^{-4}$, uncorrected), as described by Epstein and Kanwisher (1998). In addition to our functionally defined ROIs, we also defined an additional control ROI using a published “parcel” which identifies the anatomical regions corresponding to dorsal V1 (Wang et al. 2015). Within each ROI, we then calculated the magnitude of response (percent signal change) to each of our 4 experimental conditions (walking, crawling, flying, and scrambled) relative to fixation baseline using data from the Experimental Runs. A 2 (hemisphere: left, right) \times 4 (perspective: walking, crawling, flying, scrambled) repeated-measures ANOVA with a Greenhouse–Geisser correction for each scene-selective ROI was conducted. We found no significant hemisphere \times perspective interaction in OPA ($P = 0.39$), PPA ($P = 0.10$), or RSC ($P = 0.61$). Thus, both hemispheres were collapsed for further analyses.

In addition to the ROI analysis described above, we also performed a group-level analysis to explore responses to the experimental conditions across the entire slice prescription. This analysis was conducted using the same parameters used in the ROI analysis, with the exception that the experimental data were spatially smoothed with a 4-mm kernel and were registered to the standard stereotaxic (MNI) space. We then used a conjunction contrast (Nichols et al. 2005) (i.e., crawling > walking and crawling > flying and crawling > scrambled) to identify which voxels responded significantly more to crawling than all other perspectives (i.e. walking, flying, and scrambled). In other words, we only identified voxels that responded significantly more to crawling than walking, crawling and flying, “and” crawling and scrambled. Finally, conservatively, we chose the minimum *t*-statistic across these significant comparisons for visualization. The resulting statistical maps were thresholded at $P < 0.01$ (uncorrected).

Nonparametric permutation ANOVA

To further examine differences in response (or lack thereof) to the crawling, flying, and scrambled videos, we used a nonparametric permutation ANOVA (see Anderson 2001). In this analysis, we first conducted a traditional 3-level (perspective: crawling, flying, scrambled) repeated-measures ANOVA in OPA, resulting in the true *F*-statistic from our data. Then, we randomly shuffled the labels for mean percent signal change within each ROI for crawling, flying, and scrambled independently within each participant and conducted a new 3-level repeated-measures ANOVA on the permuted data resulting in a new *F*-statistic. Permutations were done within ROIs within participants and not across ROIs or participants. This process was repeated for 10,000 times, resulting in a null distribution of 10,000 *F*-values from the random permutations of our data. Then, to calculate a *P*-value (the probability

of obtaining the true *F*-value compared to the other values in our null distribution), we divided the number of *F*-values in our null distribution that were greater than the true *F*-statistic by the total number of *F*-values in our null distribution.

Video motion analysis

Motion optic flow for each video was estimated using the Farneback algorithm in the MATLAB Computer Vision Toolbox (number of pyramid layers=4, image scale=0.5, iterations per pyramid layer=3, pixel neighborhood size=7, and averaging filter size=15; Farneback 2003). The resulting estimate of flow between each frame was separated into its horizontal and vertical components (i.e. horizontal and vertical motions). The absolute values of these estimates (regardless of whether motion was right/left or up/down) were averaged across the entire video, resulting in the amount of horizontal and vertical motions in each video.

Results

OPA responded significantly more to the videos from a walking perspective than from a crawling, flying, or scrambled perspective (Fig. 2A). Indeed, a 4-level (perspective: walking, crawling, flying, scrambled) repeated-measures ANOVA revealed a significant main effect of perspective ($F(1.42, 19.81) = 5.39$, $P = 0.02$, $\eta^2_p = 0.28$), with a significantly greater response to the walking videos compared to either the crawling (main effect contrast, $P = 0.01$, $d = 0.66$), flying (main effect contrast, $P = 0.01$, $d = 0.68$), or scrambled ones (main effect contrast, $P < 0.001$, $d = 1.02$). This effect was robust; the response in OPA to the walking videos was numerically greater than both the crawling and flying videos for all 15 participants and was numerically greater than the scrambled videos for 12 of the 15 participants. By contrast, there was no significant difference between the crawling and flying videos (main effect contrast, $P = 0.95$, $d = 0.02$) or between the crawling and scrambled videos (main effect contrast, $P = 0.18$, $d = 0.35$). Taken together then, these results show that OPA is specialized for walking, not crawling through our local environment, and suggest that crawling may be processed by an entirely different neural system altogether (discussed later).

However, the similar response in OPA to the crawling, flying, and scrambled videos is essentially a null effect. Thus, does OPA really not respond to visual information about crawling? To directly address this question, we conducted 2 additional analyses.

First, we conducted a nonparametric ANOVA (Anderson 2001; see Materials and methods) in which we generated a null *F*-statistic distribution by shuffling data labels independently within each participant and then conducted a 3-level repeated-measures ANOVA on OPA’s response to the crawling, flying, and scrambled videos. These permutations were done for 10,000 times and resulted in a null *F*-statistic distribution which, compared to the true *F*-statistic of the 3-level (perspective: crawling, flying, scrambled) repeated-measures ANOVA ($F(1.22, 17.08) = 0.90$, $P = 0.38$, $\eta^2_p = 0.06$), again revealed no significant difference between OPA’s response to the crawling, flying, and scrambled videos ($P = 0.41$). Second, we conducted a 3-level Bayesian repeated-measures ANOVA, which resulted in a Bayes factor (BF_{10}) of 0.298 (Bayes factor of < 0.33 supports the null hypothesis). This analysis provides support for the null hypothesis of no difference in OPA’s response to crawling, flying, or scrambled perspectives. Thus, following these 2 additional analyses, our

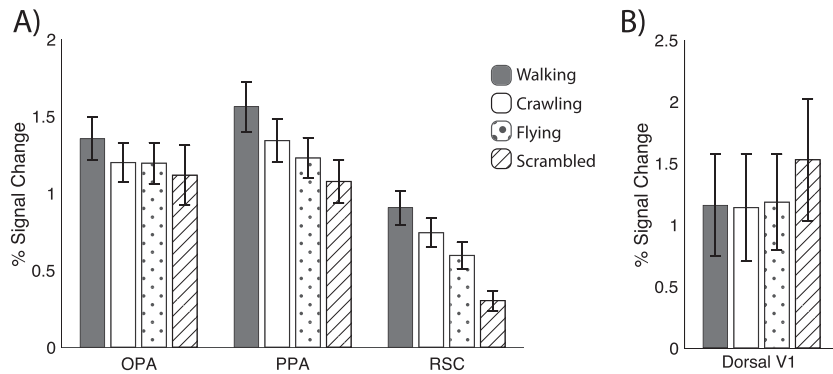


Fig. 2. Average percent signal change in all ROIs to the walking, crawling, flying, and scrambled videos. A) OPA responded significantly more to the walking videos compared to the crawling, flying, or scrambled ones and further did not respond to the crawling videos any more than to the flying or scrambled ones. Furthermore, this pattern of response is specific to OPA, and was not found in PPA or RSC. B) The pattern of response in OPA is also different from that of dorsal V1, ruling out retinotopic bias and general attention as possible explanations for our results. All error bars are \pm SEM.

findings confirm that OPA is representing visual information about walking and not crawling.

But is this pattern of response specific to OPA—consistent with its hypothesized role in visually guided navigation—or might it be a general response across all scene-selective regions, even those not involved in visually guided navigation, including the PPA and RSC? To directly test this question, we compared the response in OPA to PPA and RSC. A 3 (ROI: OPA, PPA, RSC) \times 4 (Perspective: walking, crawling, flying, scrambled) repeated-measures ANOVA revealed a significant interaction ($F(2.16, 29.62) = 9.74, P < 0.001, \eta^2_p = 0.41$) (Fig. 2A). A linear trend analysis then revealed a significant ROI \times linear trend interaction ($F(1, 14) = 7.44, P = 0.02, \eta^2_p = 0.347$) with a significant linear decrease (from crawling to flying to scrambled) in PPA ($F(1, 14) = 11.31, P = 0.005, \eta^2_p = 0.422$) and RSC ($F(1, 14) = 55.89, P < 0.001, \eta^2_p = 0.80$), but not OPA ($F(1, 14) = 0.88, P = 0.36, \eta^2_p = 0.059$), revealing a qualitatively different response in OPA compared to either PPA or RSC. Thus, the selective response to visual information about walking is specific to OPA, which is consistent with its hypothesized role in visually guided navigation.

However, recent work has shown that OPA has a retinotopic bias for information in the lower visual field (Silson et al. 2015). So, might this retinotopic bias somehow account for our OPA results? Perhaps, there is more visual information in the lower visual field in the walking videos, compared to the other perspectives, and the response in OPA simply reflects its lower visual field bias. To directly test this possibility, we examined the responses in OPA and dorsal V1 (which represents visual information in the lower visual field). If our results can be explained by more information in the lower half of the visual field, then we would “not” expect to see a difference between OPA and dorsal V1. However, a 2 (ROI: OPA, dorsal V1) \times 4 (perspective: walking, crawling, flying, scrambled) repeated-measures ANOVA revealed a significant interaction ($F(2.32, 32.49) = 15.20, P < 0.001, \eta^2_p = 0.52$) (Fig. 2A and B). Thus, our results cannot be explained by OPA’s retinotopic bias. Also, note, this significant interaction rules out the possibility that participants were simply paying more attention to the walking videos than to the other videos. If this “attentional” account was true, then both OPA and dorsal V1 would respond significantly more to the walking videos than all other perspectives. However, the response in dorsal V1 was actually significantly greater for the scrambled videos compared to all the other perspectives (perhaps, due to the additional “edges” induced by scrambling)—the complete opposite pattern of OPA.

But why does OPA respond more to videos from a walking perspective than from a crawling perspective? In other words, what differences in visual information between our walking and crawling videos are driving OPA’s selective response to the walking videos? One possibility could be the different kinds of motion between the walking and crawling videos. For example, perhaps there is more vertical motion in the crawling videos (as a result of needing to pick the head up to see where one is going) than in the walking videos (Kretch et al. 2014). To directly address this question, using Farneback’s motion estimation algorithm (see Materials and methods for more detail), we then compared the average amount of horizontal and vertical motions in each walking and crawling video. Indeed, a 2 (perspective: walking, crawling) \times 2 (motion direction: horizontal, vertical) mixed ANOVA revealed a significant interaction between perspective and direction of motion ($F(1, 22) = 6.18, P = 0.02, \eta^2_p = 0.219$; Fig. 3), with post hoc contrasts revealing more vertical motion (compared to horizontal motion) in the crawling ($t(22) = 2.88, P = 0.01$) but not walking videos ($t(22) = 0.634, P = 0.53$).

Finally, if OPA indeed responds only to visual information from a walking perspective, and not crawling, then what system supports crawling? To explore this question, we conducted a group-level whole-brain analysis to find regions which respond more to the crawling videos than to the walking, flying, and scrambled ones using a conjunction contrast (i.e. crawling > walking and crawling > flying and crawling > scrambled, all $P_s < 0.01$ uncorrected). We found bilateral regions in the inferior parietal lobule in addition to other regions in the bilateral superior parietal lobule extending into premotor cortex, which responded more to the crawling videos than all other perspectives (i.e. walking, flying, and scrambled; Fig. 4). Future work is needed to directly test the role of these regions in moving about the environment via crawling.

Discussion

In this study, we investigated the precise role of OPA in visually guided navigation, and, more specifically, asked whether OPA represents visual information from the 2 perspectives by which humans move about their local environment (i.e. crawling and walking), or instead represents visual information about walking only and not crawling. We found that OPA responded more to the walking videos than to the crawling, flying, and scrambled ones; and, moreover, did not differentiate between the crawling, flying,

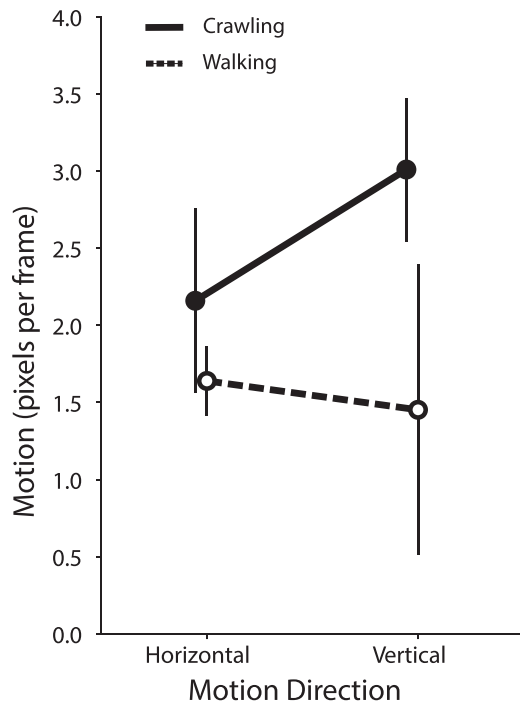


Fig. 3. Average horizontal and vertical motions in walking and crawling videos. There was significantly more vertical motion (compared to horizontal motion), in the crawling, not walking, videos. All error bars are \pm SEM.

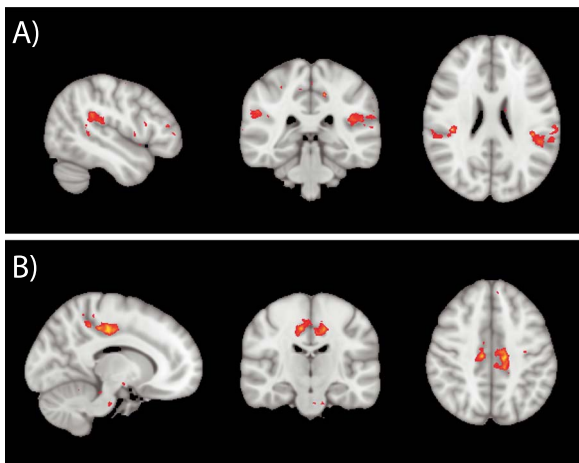


Fig. 4. A group conjunction contrast showing regions which responded significantly more to the crawling videos than to the walking and flying and scrambled ones ($P < 0.01$, uncorrected). These regions include A) bilateral inferior parietal lobule (MNI coordinates: 137, 92, and 96) and B) bilateral superior parietal lobule extending into premotor cortex (MNI coordinates: 102, 107, and 115).

and scrambled videos—demonstrating that OPA represents visual information about walking only and not crawling. Critically, these results were not due to general scene selectivity, retinotopic bias, or attentional differences between the different video perspectives.

We further investigated what differences in visual information between our walking and crawling videos may have driven OPA's selective response to the walking videos. A video motion analysis revealed that there are different kinds of motion between walking and crawling videos (with crawling videos having more vertical motion). However, there are also likely other differences between

these 2 perspectives which drove OPA's selective response to the walking videos. For example, OPA has been shown to represent egocentric distance (near/far) information (Persichetti and Dilks 2016), and perhaps there was more information about egocentric distance in our walking videos compared to crawling videos (due to the head being slightly angled toward the ground when crawling). Future work is needed to explore the entirety of visual information differences (besides the kind of motion) that may drive OPA's selective response to walking.

The finding that OPA does not respond any more to the crawling videos than to the scrambled ones supports the conclusion that OPA does not represent visual information about crawling, but it also raises an intriguing question: Why does OPA respond to these scrambled videos at all when considering that they are actually not navigable? One possibility is that OPA responded to our scrambled videos because OPA was processing information about the local elements of the scenes (Kamps, Julian, et al. 2016), which were likely present in our scrambled videos. However, while the scrambled videos may have contained some of the information needed for visually guided navigation, and hence why OPA responded significantly more to the scrambled videos compared to a fixation baseline, the fact that OPA responded significantly more to the walking videos (which actually mimic the visual experience of navigating), compared to all of the other videos, indicates that OPA is specialized for walking.

Finally, why is OPA only representing visual information about walking, not crawling, given that we move about our local visual environments before we walk, beginning with crawling as an infant? One possibility is that OPA has never supported crawling throughout development. Consistent with this idea, recent work has found that OPA does not even represent first-person perspective motion information in children at 5 years of age and rather only emerges at around 8 years of age (Kamps et al. 2020). Considering then that OPA is not fully functioning until so late in development, it seems likely that another system (other than OPA) supports our ability to crawl around the environment. Indeed, the results of our group-level analysis suggests that both the inferior and superior parietal lobules may be involved in processing visual information supporting our ability to crawl about our local environments, but future work is needed to directly investigate this possibility.

Conclusion

In conclusion, we found that OPA responds only to visual information from 1 perspective by which humans move through their local environments (i.e. walking), and not from perspectives by which humans do not (i.e. flying and scrambled), supporting the hypothesis that OPA is involved in visually guided navigation. However, we also found that OPA does not respond to crawling videos any more than to either flying or scrambled ones, suggesting that OPA does not represent visual information about crawling, which is consistent with the hypothesis that OPA undergoes protracted development. Finally, our results suggest that OPA may have never supported crawling and that visually guided navigation may undergo a discontinuous developmental trajectory; however, future work is needed to directly investigate these possibilities.

CRedit authors statement

Christopher M. Jones (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization,

Writing—original draft, Writing—review & editing), Joshua Byland (Formal analysis, Software, Validation), and Daniel D. Dilks (Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Writing—original draft, Writing—review & editing)

Acknowledgments

We would like to thank Yaelan Jung for assistance with data collection and helpful comments on the manuscript. We would also like to thank the Facility for Education and Research in Neuroscience (FERN) Imaging Center in the Department of Psychology, Emory University, Atlanta, GA.

Funding

This work was supported by a grant from the National Eye Institute—R01 EY29724 (DDD).

Conflict of interest statement: None declared.

Data availability

All datasets generated in this study will be available to readers upon reasonable request.

References

- Adolph KE, Vereijken B, Denny MA. Learning to crawl. *Child Dev.* 1998;69(5):1299–1312.
- Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral ecology.* 2001;26(1):32–46.
- Bonner MF, Epstein RA. Coding of navigational affordances in the human visual system. *Proc Natl Acad Sci.* 2017;114(18):4793–4798.
- Cheng A, Walther DB, Park S, Dilks DD. Concavity as a diagnostic feature of visual scenes. *NeuroImage.* 2021;232:117920.
- Cox RW, Jesmanowicz A. Real-time 3D image registration for functional MRI. *Magn Reson Med.* 1999;42(6):1014–1018.
- Persichetti AS, Dilks DD. Perceived egocentric distance sensitivity and invariance across scene-selective cortex. *Cortex.* 2016;77:155–163.
- Dilks DD, Julian JB, Kubilius J, Spelke ES, Kanwisher N. Mirror-image sensitivity and invariance in object and scene processing pathways. *J Neurosci.* 2011;31(31):11305–11312.
- Dilks DD, Julian JB, Paunov AM, Kanwisher N. The occipital place area (OPA) is causally and selectively involved in scene perception. *J Neurosci.* 2013;33:1331–1336.
- Dilks DD, Kamps FS, Persichetti AS. Three cortical scene systems and their development. *Trends Cogn Sci.* 2022;26(2):117–127.
- Dillon MR, Persichetti AS, Spelke ES, Dilks DD. Places in the brain: bridging layout and object geometry in scene-selective cortex. *Cereb Cortex.* 2018;28(7):2365–2374.
- Epstein R, Kanwisher N. A cortical representation of the local visual environment. *Nature.* 1998;392:598–601.
- Farneäck G. Two-frame motion estimation based on polynomial expansion. In: *Scandinavian conference on image analysis.* Berlin, Heidelberg: Springer; 2003. pp. 363–370.
- Henriksson L, Mur M, Kriegeskorte N. Rapid invariant encoding of scene layout in human OPA. *Neuron.* 2019;103(1):161–171.
- Kamps FS, Julian JB, Kubilius J, Kanwisher N, Dilks DD. The occipital place area represents the local elements of scenes. *NeuroImage.* 2016;132:417–424.
- Kamps FS, Lall V, Dilks DD. The occipital place area represents first-person perspective motion information through scenes. *Cortex.* 2016;83:17–26.
- Kamps FS, Pincus JE, Radwan SF, Wahab S, Dilks DD. Late development of navigationally relevant motion processing in the occipital place area. *Curr Biol.* 2020;30(3):544–550.
- Kretch KS, Franchak JM, Adolph KE. Crawling and walking infants see the world differently. *Child Dev.* 2014;85(4):1503–1518.
- Nichols T, Brett M, Andersson J, Wager T, Poline JB. Valid conjunction inference with the minimum statistic. *NeuroImage.* 2005;25(3):653–660.
- Park J, Park S. Coding of navigational distance and functional constraint of boundaries in the human scene-selective cortex. *J Neurosci.* 2020;40(18):3621–3630.
- Persichetti AS, Dilks DD. Dissociable neural systems for recognizing places and navigating through them. *J Neurosci.* 2018;38:10295–10304.
- Silson EH, Chan AWY, Reynolds RC, Kravitz DJ, Baker CI. A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *J Neurosci.* 2015;35(34):11921–11935.
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, de Luca M, Drobnjak I, Flitney DE, et al. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage.* 2004;23:S208–S219.
- Wang L, Mruczek RE, Arcaro MJ, Kastner S. Probabilistic maps of visual topography in human cortex. *Cereb Cortex.* 2015;25(10):3911–3931.