Distribution Agreement

Signature:

________________________________ _______________
Christopher Jones Date

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

By

Christopher Jones
B.S., University of Utah

Psychology

Daniel D. Dilks, Ph.D. ___________________
Advisor

Philip Kragel, Ph.D.______________________
Committee Member

Lynne Nygaard, Ph.D.____________________
Committee Member

Accepted:

____
Kimberly Jacob Arriola, PhD, MPH
Dean of the James T. Laney School of Graduate Studies

____
Date

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

By

Christopher M. Jones
B.S.

Advisor: Daniel D. Dilks. Ph.D.

An abstract of
A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Arts
in Psychology
2022

Abstract

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

By Christopher M. Jones

It has recently been hypothesized that the occipital place area (OPA) – a scene-selective region in adult humans – supports “visually-guided navigation” (i.e., finding our way through the local visual environment, avoiding boundaries and obstacles). Here we directly test this hypothesis using functional magnetic resonance imaging (fMRI) in adult humans. Specifically, we measured the responses in OPA to videos depicting the actual first-person visual experience of moving through local environments – from either a “walking” or “crawling” perspective, as well as two control conditions: “flying” and “scrambled”. If OPA indeed supports visually-guided navigation, then it will respond only to visual information from perspectives by which humans actually move through their local environments (i.e., beginning with crawling in infancy followed by walking), and not from perspectives by which humans do not (i.e., flying and scrambled). Consistent with our hypothesis, we found that that OPA responded significantly more to the walking videos compared to the flying and scrambled ones. Surprisingly, however, OPA also responded more to the walking videos than to the crawling videos, and, moreover, responded similarly to the crawling, flying, and scrambled videos. Taken together, these results i) reveal that OPA is involved in visually-guided navigation, but only from a walking perspective, not from a crawling perspective, ii) suggest that crawling is processed by an entirely different neural system, and iii) raise intriguing questions for how OPA develops; namely, that OPA may have never supported crawling during development, consistent with the hypothesis that OPA undergoes protracted development.

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

By

Christopher M. Jones
B.S.

Advisor: Daniel D. Dilks. Ph.D.

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Arts
in Psychology
2022

CONTENTS

Introduction

Finding our way through 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 occipital place area (OPA) (Dilks et al., 2013) – is selectively involved in visually-guided navigation (Dilks et al., 2022). Indeed, several fMRI studies found that the OPA represents at least four kinds of information relevant for visually-guided navigation: i) "sense" (left/right) information (Dilks et al., 2011); ii) egocentric distance (near/far) information (Persichetti & Dilks, 2016); iii) local scene elements ("parts"), including boundaries (e.g., walls) and/or obstacles (e.g., furniture) that constrain how one can move about the local visual environment (Henriksson et al., 2019; Park & Park, 2020; Kamps et al., 2016a; Cheng et al., 2021; Dillon et al., 2018), and iv) possible routes through a local scene (Persichetti & Dilks, 2018; Bonner & Epstein, 2017).

Perhaps even more comprehensive though is another fMRI study (Kamps et al., 2016b) 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 provide an even stronger test of the hypothesis that OPA supports visually-guided navigation by comparing the response in OPA to videos depicting the actual first-person visual experience from the two perspectives by which humans actively move about the local environment (i.e., “walking” and “crawling”) versus two other videos by which humans do not move about the local environment (i.e., a top-down arial perspective, or “flying”) or cannot move about the local environment (i.e., “scrambled” versions of the walking videos) (Figure 1). If OPA is indeed involved in visually-guided navigation, then it will respond significantly more to both the walking and crawling videos, compared to the flying and scrambled ones.

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. For a follow-up study, three additional participants (ages 19 - 26, mean age 22.3; 1 female, 2 males). These three participants also had normal or corrected to normal vision, reported no history of neurological conditions, gave informed consent, and were compensated for their participation. All procedures were approved by the Emory University Institutional Review Board.

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 two control conditions: "flying" and "scrambled" (Experimental Runs). For both the Localizer and Experimental Runs, participants performed a one-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 seconds long, and consisted of 4 blocks per stimulus category. For each run, the order of the first eight blocks was pseudorandomized, and the order of the remaining eight blocks was the palindrome of the first eight. Each block contained 20 images from the same category for a total of 16 second blocks. Each image was presented for 300ms, followed by a 500 ms interstimulus interval, and subtended 8 by 8° of visual angle. We also included five 16-second fixation blocks: one at the beginning, three in the middle interleaved between each palindrome, and one at the end of each run.

For the Experimental runs, we made a total of 12 3-second video clips for each of our experimental conditions ("walking", "crawling", "flying", "scrambled"). These videos were filmed using a GoPro camera. For the walking videos, the videos were taken while one of the authors (J.B.) – with the camera attached to his forehead – walked through 12 different places (e.g., a backyard, a parking lot, a hallway). For the crawling videos, the videos were taken while J.B – 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 10 feet in the air, facing down at the ground rather than facing out as if walking, while J.B. walked through the same 12 places again. Note: Our goal was not to actually simulate flying, but rather show a video perspective by which humans do not navigate. Finally, for our scrambled videos, we divided our walking videos into a 9 x 9 grid, and permutated this grid to scramble the video. All of these video clips subtended approximately 15.7 x 20.7° visual angle. All participants completed 8 experimental runs, however in two 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 seconds long, and consisted of 4 blocks of 6 videos for each perspective condition. For each run, the order of the first eight blocks was pseudorandomized, and the order of the remaining eight blocks was the palindrome of the first eight. Before each block, a word was presented for 2 seconds 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 seconds followed by a 300ms inter-stimulus interval between each video, resulting in a total block length of 21.8 seconds. We also included five 19.8-second fixation blocks: one at the beginning, three in the middle interleaved between each palindrome, and one at the end of each run.

The additional 3 participants for the follow-up study each completed 4 experimental runs in which they viewed 2 blocks of 6 videos (the same videos from the main experiment) for each experimental condition (walking, crawling, flying, and scrambled). We also included three fixation blocks (each lasting 12 seconds): one at the beginning, one in the middle of the palindrome, and one at the end. There was no task and participants were asked to imagine themselves navigating from each perspective.

fMRI scanning. All scanning was performed on a 3T Siemen 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 x 3 x 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 FreesSurfer Functional Analysis Stream, while statistical analyses were conducted using JASP (Love et al., 2019). ROI analyses were done using the FS-FAST ROI toolbox. Before statistical analysis, images were motion corrected (Cox & Jesmanowicz, 1999). Data were then detrended and fit using a double gamma function. Localizer data, but not experimental data, were spatially smoothed with a 5mm kernel. After preprocessing, 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 & Kanwisher (1998). In addition to our functionally-defined ROIs, we also defined two additional control ROIs using two published “parcels” which identify the anatomical regions corresponding to motion area MT (Julian et al., 2012), as well as 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, scrambled) using data from the experimental runs. A 2 (Hemisphere: Left, Right) x 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 x perspective interaction in OPA ($p = .39$), PPA ($p = .10$), or RSC ($p = .61$). Thus, both hemispheres were collapsed for further analyses. In our additional 3 subjects, OPA was defined using a published parcel (Julian et al., 2012).

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 as were used in the ROI analysis, with the exceptions that the experimental data were spatially smoothed with a 4-mm kernel, and registered to standard stereotaxic (MNI) space. We then used a conjunction contrast (Nichols et al., 2004) in which we took the minimum t-statistic from the following contrasts: crawling > walking and crawling > flying and crawling > scrambled. The resulting statistical maps were then thresholded at $p <.05$ (uncorrected) to identify which voxels responded significantly more to crawling than walking, flying, and scrambled.

Non-parametric permutation ANOVA. In order to further examine differences in response (or lack thereof) to the crawling, flying, and scrambled videos, we used a non-parametric 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 crawling, flying, and scrambled independently within each participant and conducted a new 3 level repeated-measures ANOVA on the permutated data resulting in a new F-statistic. This process was repeated 10,000 times resulting in a null distribution of 10,000 F-values from random permutations of our data. Then, in order 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.

Results

Consistent with our prediction, OPA responded significantly more to videos from a walking perspective than from a flying or scrambled perspective. Contradictory to our prediction, however, OPA also responded significantly more to videos from a walking perspective than from a crawling perspective (Figure 2A). Indeed, a 4 level (Perspective: walking, crawling, flying, scrambled) repeated-measures ANOVA revealed a significant main effect of perspective ($F(1.42, 19.81) = 5.39$, $p = 0.02$, $\eta^2_p = .28$), with a significantly greater response to the walking videos compared to either the crawling (main effect contrast, $p = .01$, $d = .66$), flying (main effect contrast, $p = .01$, $d = .68$), or scrambled ones (main effect contrast, $p < .001$, $d = 1.02$). This effect was robust; the response in OPA to the walking videos was numerically greater than both the crawling and flying videos in all 15 participants, and numerically greater than the scrambled videos in 12 of the 15 participants. By contrast, there was no significant difference between the crawling and flying videos (main effect contrast, $p = .95$, $d = .02$) nor between the crawling and scrambled videos (main effect contrast, $p = .18$, $d = .35$). Taken together then, these results support the hypothesized role of OPA in visually-guided navigation, and suggest that crawling may not be a mode of visually-guided navigation at all.

Given the surprising finding that OPA responded significantly more to the walking than crawling videos, and similarly to the crawling, flying, and scrambled videos, we next examined whether this effect replicated in another group of participants all together, using a slightly different paradigm (see Methods). In this new group of three participants, we again found that OPA responded more to the walking than crawling, flying, and scrambled videos, and similarly to the crawling, flying, and scrambled videos (Figure 2B). However, while a 4-level repeated measures ANOVA did not reveal a significant main effect of perspective ($F(3, 6) = 2.21$, $p = .18$) – no doubt given the limited sample size – we did find a large effect size ($\eta^2_p = .52$), indicating that this effect likely would be significant with a larger sample size.

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 two additional analyses.
First, we conducted a non-parametric ANOVA (Anderson, 2001; see 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 crawling, flying, and scrambled videos. These permutations were done ten thousand 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 ($F(1.22, 17.08) = .90$, $p = .38$, $\eta^2_p = .06$), again revealed no significant difference between OPA's response to the crawling,

flying, and scrambled videos ($p = .41$). Second, we conducted a 3 level Bayesian repeated-measures ANOVA which resulted in a Bayes factor of 0.30, (Bayes factor of <0.33 supports the null hypothesis, Bayes factor > 3 supports the alternative 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 two additional analyses, our findings confirm that OPA is representing visual information about walking, not crawling.

But is this pattern of response really 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-guiding navigation, including the parahippocampal place area (PPA) and retrosplenial complex (RSC)? To test this question, we directly compared the response in OPA to PPA and RSC. A 3 (ROI: OPA,PPA, RSC) x 4 (Perspective: walking, crawling, flying, scrambled) repeated-measures ANOVA revealed a significant interaction ($F(2.16, 29.62) = 9.74$, $p < .001$, $\eta^2_p = .41$) (Figure 2A). A linear trend analysis then revealed a significant ROI x linear trend interaction ($F(1, 14) = 7.44$, $p = .016$, $\eta^2_p = .347$) with a significant linear decrease (from crawling to flying to scrambled) in PPA ($F(1, 14) = 11.31$, $p = .005$, $\eta^2_p = .422$) and RSC($F(1, 14) = 55.89$, $p < .001$, $\eta^2_p = .80$), but not OPA ($F(1, 14) = .88$, $p = .364$, $\eta^2_p = .059$), 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, 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) x 4 (Perspective: walking, crawling, flying, scrambled) repeated-measures ANOVA revealed a significant interaction (F(2.32, 32.49) = 15.20, $p < .001$, $\eta^2_p = .52$) (Figure 2A and 2C). 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 were true, then both OPA and dorsal V1would respond significantly more to the walking videos than all other perspectives. However, the response in dorsal V1 was actually significantly greater for scrambled videos compared to all of the other perspectives; the complete opposite of OPA.

Next, we considered whether motion might explain our results. Maybe there is simply more motion in the walking videos compared to crawling, flying, or scrambled ones. To rule out this possibility, we compared the responses of MT (a general motion selective region) to OPA. Critically, a 2 (ROI: OPA, MT) x 4 (Perspective: walking, crawling, flying, scrambled) repeated-measures ANOVA revealed a significant interaction (F(1.55,21.77) = 19.99, $p <.001$, $\eta^2_p = .59$) (Figure 2A and 2C). Thus, the selective response in OPA to the walking videos cannot be explained by differential motion processing in OPA.

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 compared to the walking, flying, and scrambled ones (Figure 3). Future work is needed to directly test the role of these regions in moving about the environment via crawling.

Discussion

In this study we directly tested the role of OPA in visually-guided navigation by asking whether OPA responds more to visual information from perspectives by which humans actually move about their local environments (i.e., by walking and crawling), compared to perspectives by which they do not (i.e., flying and scrambled). Consistent with the hypothesis that OPA is involved in visually-guided navigation, we indeed found that OPA responded more to the walking videos than to the flying and scrambled ones. However, contrary to our predictions, OPA also responded more to the walking videos than to the crawling ones, and surprisingly did not differentiate between the crawling, flying, and scrambled videos. Thus, this finding indicates that OPA represents visual information about walking, not crawling. Critically, we found that this pattern of results is unique to OPA, hence indicative of its role in visually-guided navigation, and not due to general scene selectivity, retinotopic bias, attention, or general motion differences between different video perspectives.

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 why does OPA respond to these scrambled videos at all, 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 et al., 2016a), which were likely present in our scrambled videos. However, while the scrambled videos (or even the crawling and flying videos) may have contained some of the information needed for visually-guided navigation, and hence why OPA responded above baseline (fixation) to this condition, the fact that OPA responded the most to walking videos (which mimic the visual experience of actually navigating) indicates that OPA's primary function is visually-guided navigation, but only from a walking perspective.

So, why is OPA only representing visual information about walking, and not crawling, given that we move about our local 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 isn't 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 the inferior parietal lobule may be involved in the ability to crawl, but future work is needed to directly investigate this possibility.

Finally, our data — which suggest that walking and crawling are processed by different neural systems — challenge the common thought that crawling is a direct developmental milestone towards our ability to walk (and navigate) using our legs. Indeed, our results are consistent with other developmental data dissociating walking and crawling as qualitatively different modes of locomotion. For example, Gibson et al., (1987) found that infants who had recently begun walking were able to discriminate navigational affordances of different surfaces (i.e., a firm surface suitable for walking versus an unstable waterbed), while crawling infants were not. Relatedly, Adolph (1997) trained crawling infants on the locomotive affordances about the slope of a “hill”, but slope information about this “hill” learned while crawling did not transfer once the infants started walking, and they had to learn slope information all over again. Finally, Clearfield (2004) found that knowledge about landmarks also does not transfer between crawling and walking, and thus is tightly coupled with the mode of locomotion. Collectively, these studies suggesting that crawling and walking are dissociable systems is consistent with our findings that OPA is selective for the visual information from a walking perspective, but not crawling.

In conclusion, we found that OPA responds only to visual information from one 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, 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.

References

Adolph, K. E., Bertenthal, B. I., Boker, S. M., Goldfield, E. C., & Gibson, E. J. (1997). Learning in the development of infant locomotion. *Monographs of the society for research in child development*, i-162

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral ecology*, *26*(1), 32-46.

Bonner, M. F., & Epstein, R. A. (2017). Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences*, *114*(18), 4793-4798.

Cheng, A., Walther, D. B., Park, S., & Dilks, D. D. (2021). Concavity as a diagnostic feature of visual scenes. *NeuroImage*, *232*, 117920.

Clearfield, M. W. (2004). The role of crawling and walking experience in infant spatial memory. *Journal of experimental child psychology*, *89*(3), 214-241.

Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *42*(6), 1014-1018

Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-image sensitivity and invariance in object and scene processing pathways. *Journal of Neuroscience*, *31*(31), 11305-11312.

Dilks, D. D., Kamps, F. S., & Persichetti, A. S. (2021). Three cortical scene systems and their development. *Trends in cognitive sciences*.

Dilks, D.D., Julian, J.B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area (OPA) is causally and selectively involved in scene perception. *Journal of Neuroscience*, 33, 1331-1336. https://doi.org/10.1523/JNEUROSCI.4081-12.2013

Dillon, M. R., Persichetti, A. S., Spelke, E. S., & Dilks, D. D. (2018). Places in the brain: Bridging layout and object geometry in scene-selective cortex. *Cerebral cortex*, *28*(7), 2365-2374.

Epstein, R., Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature* 392, 598–601.

Gibson, E. J., Riccio, G., Schmuckler, M. A., Stoffregen, T. A., Rosenberg, D., & Taormina, J. (1987). Detection of the traversability of surfaces by crawling and walking infants. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(4), 533

Henriksson, L., Mur, M., & Kriegeskorte, N. (2019). Rapid invariant encoding of scene layout in human OPA. *Neuron*, *103*(1), 161-171.

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *Neuroimage*, *60*(4), 2357-2364.

Kamps F.S., Lall, V. Dilks, D.D. (2016). The occipital place area represents first-person perspective motion information through scenes, Cortex, Volume 83, Pages 17-26.

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. *Neuroimage*, *132*, 417-424.

Kamps, F. S., Pincus, J. E., Radwan, S. F., Wahab, S., & Dilks, D. D. (2020). Late Development of Navigationally Relevant Motion Processing in the Occipital Place Area. *Current Biology*, *30*(3), 544-550.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., ... & Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software, 88,* 1-17.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*(3), 653-660.

Park, J., & Park, S. (2020). Coding of navigational distance and functional constraint of boundaries in the human scene-selective cortex. *Journal of Neuroscience*, *40*(18), 3621-3630.

Persichetti, A.S, & Dilks, D.D. (2016). Perceived egocentric distance sensitivity and invariance across scene-selective cortex, *Cortex*, Volume 77, Pages 155-163.

Persichetti, A.S., & Dilks, D.D. (2018). Dissociable neural systems for recognizing places and navigating through them. *Journal of Neuroscience*, 38, 10295-10304.

Silson, E. H., Chan, A. W. Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *Journal of Neuroscience*, *35*(34), 11921-11935.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., ... & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*, *23*, S208-S219.

Wang, L., Mruczek, R. E., Arcaro, M. J., & Kastner, S. (2015). Probabilistic maps of visual topography in human cortex. *Cerebral cortex*, *25*(10), 3911-3931.

Figures

Figure 1. Example frames from the “walking”, “crawling”, “flying”, and “scrambled” videos.

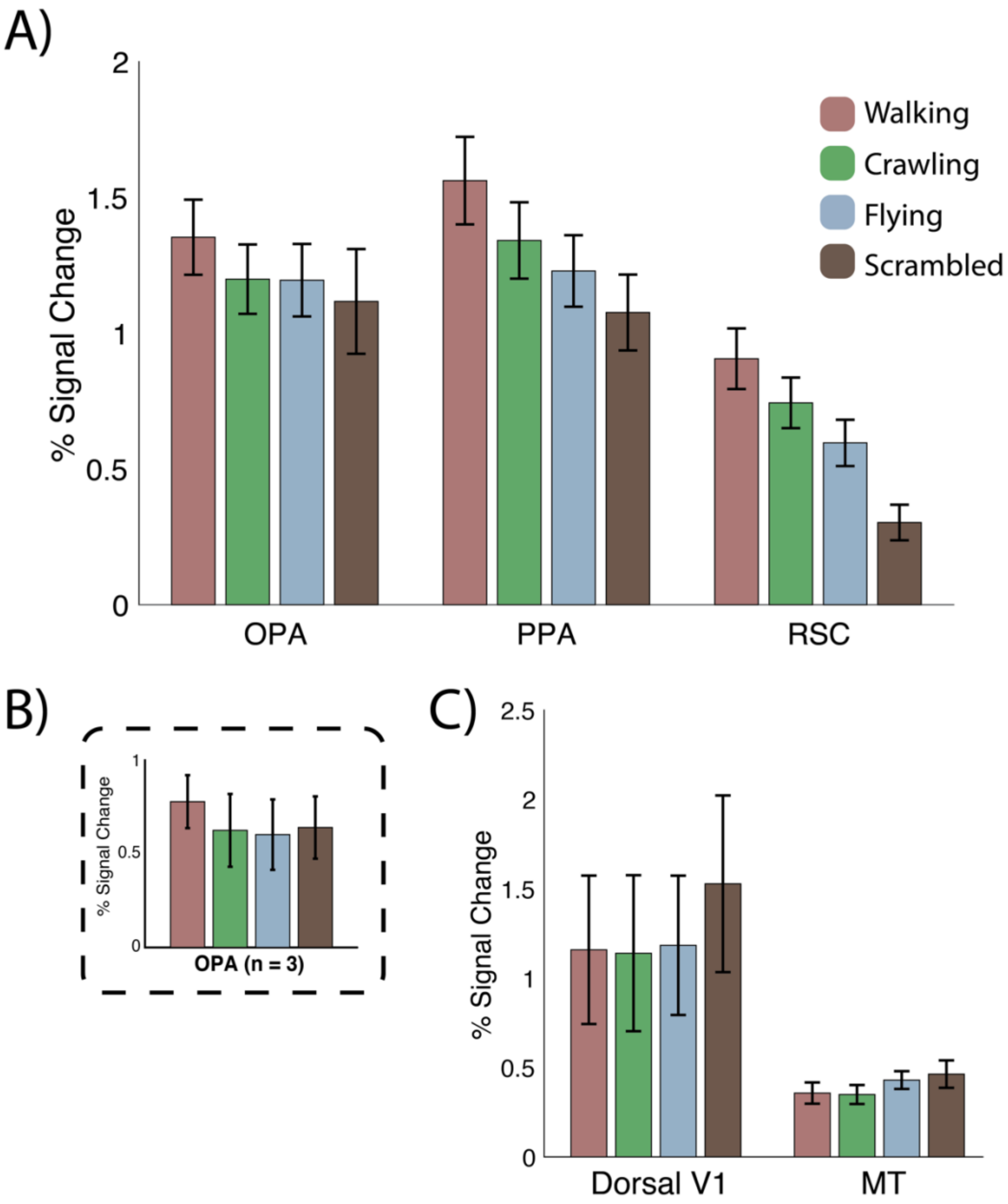

Figure 2. Average percent signal change in all ROIs to the walking, crawling, flying, and scrambled videos. (A) OPA responded significantly more to the walking videos compared to the crawling, flying, or scrambled ones, and further did not respond to the crawling videos any more than to the flying or scrambled ones. Furthermore, this pattern of response is specific to OPA, and was not found in PPA or RSC. (B) The response in OPA to the walking, crawling, flying, and scrambled videos from an additional sample of 3 participants. Note that the pattern of response is identical to that of OPA in the main experiment. (C) The pattern of response in OPA is also different than that of both dorsal V1 and MT, ruling out, OPA's retinotopic bias, general motion, and general attention as possible explanations for our results. All error bars are ± SEM.

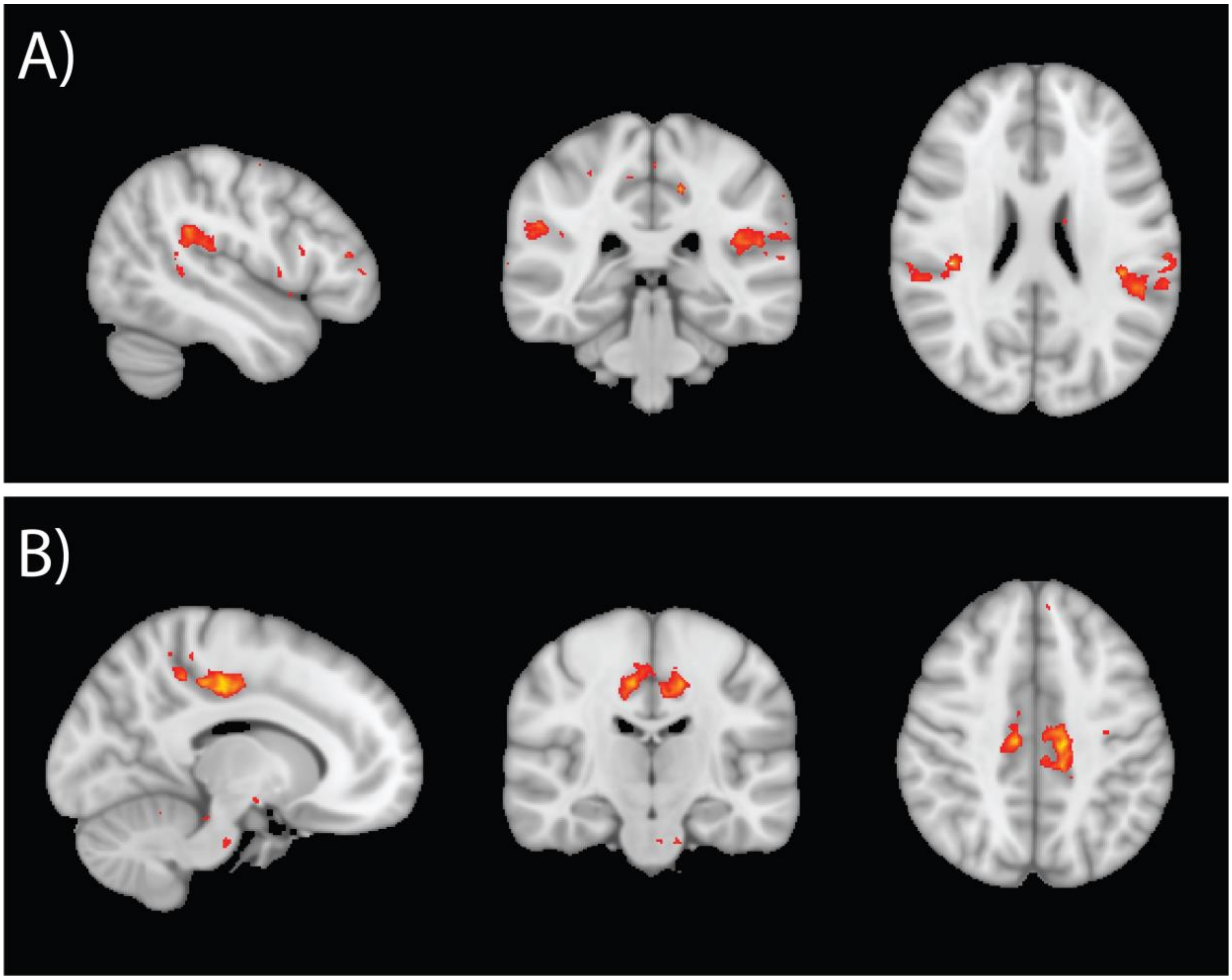

Figure 3. A group conjunction contrast showing regions which responded significantly more to the crawling videos than to the walking and flying and scrambled ones (p < .01, uncorrected). These regions include (A) bilateral inferior parietal lobule (MNI coordinates: 137, 92, 96) and, (B) bilateral superior parietal lobule extending into premotor cortex (MNI coordinates: 102, 107, 115).