Title
Modelling navigation representations during naturalistic driving

Author
Zhang, Tianjiao

Publication Date
2021

Peer reviewed|Thesis/dissertation

Modeling navigation representations during naturalistic driving

by

Tianjiao Zhang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy

with University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jack Gallant, Chair

Professor Srikantan Nagarajan

Professor Michael Silver

Professor Michael Yartsev

Fall 2021

Modeling navigation representations during naturalistic driving

Abstract

Modeling navigation representations during naturalistic driving

by

Tianjiao Zhang

Doctor of Philosophy in Bioengineering

University of California, Berkeley

Professor Jack Gallant, Chair

Navigation in the natural world is a complex task that engages many cognitive systems, including vision, attention, motor control, cognitive maps, and planning. These systems recruit many brain regions that form multiple functional networks spanning the brain. Navigation has been the subject of many non-human neurophysiology studies, primarily in rodents, and less frequently in non-human primates.

In the recent decades, neuroimaging using functional magnetic resonance imaging has enabled non-invasive brain activity recordings. Neuroimaging studies are able to examine cognitive processes in healthy human subjects. Together, neurophysiology and neuroimaging studies have revealed multiple regions in the human brain that are active during navigation. However, we still know very little about how these regions accomplish the complex task of navigation, and what information is represented by each region. This dissertation describes a next-generation neuroimaging experiment that maps navigational representations across the cortex to gain insight into the complex processes underlying human navigation.

Chapter 1 reviews the current neurophysiology and neuroimaging literature on the navigation system of the brain. These current literatures suggest that the brain breaks navigation down into a set of subtasks, each with its associated brain regions. This chapter also describes several limitations in the current understanding of navigation in the human brain. Chapter 2 describes a next-generation, naturalistic neuroimaging paradigm to study human navigation. This paradigm leverages Unreal Engine 4, a modern game engine, to create a realistic and dynamic virtual world. A custom MR-compatible steering wheel and pedal set enables subjects to drive naturally in this virtual world. Chapter 3 describes an active navigation neuroimaging experiment that uses the paradigm developed in Chapter 2. Subjects actively perform a taxi driver task in a virtual world while BOLD activity is recorded from the brain. Voxelwise modeling with banded ridge regression is used to map the cortical representation of over 20,000 features across 33 feature spaces. Chapter 4 examines task-related visual semantic tuning shifts. Visual semantic tunings in the active navigation task are compared with those from a passive movie watching task. Results show that there are significant visual semantic tuning shifts between active navigation and passive movie watching. Chapter 5 describes a route progression model derived from the rodent literature. Results suggest that a topologic representation of route progression in RSC is conserved across species. Furthermore, RSC, unlike other visual navigation areas, strongly prefers the

start of routes. Finally, Chapter 6 describes a novel, state-space based method for analyzing high-dimensional brain data. This method treats the activity of the brain as a dynamical system, and finds a low-dimensional subspace in the brain's activity space that is related to the representation of task variables. As a proof-of-concept, this method is applied to a visual semantic attention task and a video game task, and recovers low-dimensional spaces related to the representation of attentional targets and game states.

Contents

Figures

Tables

Acknowledgments

A few thanks are in order.

Well. More than a few.

Many thanks are in order.

First and foremost, I would like to thank my advisor, Jack Gallant. I could not have done this work without his advice, support, and mentorship. Jack provided me the freedom to pursue a manifestly impossible project, and the scientific insight and acumen to turn it into something real.

I would also need to thank members of my dissertation and quals committees, Michael Silver, Michael Yartsev, Sri Nagarajan, and Frederic Theunissen, for giving me so much of their time, advice, and perspective. They have been instrumental in my professional development.

Much credit would need to be given to my lab mates. I have had countless fruitful discussions with Mark Lescroart about experiment design, interpretation, and science in general. He has continued to provide advice and mentorship even though he is now a professor running his own lab. James Gao created the Ur-experiment that planted the seed for my project, and single-handedly kept the computers up and running. Anwar Nunez-Elizalde provided many ideas as I built the project, and laid the mathematical framework upon which all these analyses here are based. Storm Slivkoff answered all my dumb questions when I rotated through the lab, and all my dumb questions after I joined. Sara Popham, because she was the same year I am, gave me the motivation to get all the things done, like doing quals and writing my dissertation and trying to graduate. Michael Eickenberg helped with much mathematical thinking, and was my guinea pig subject. Tom Dupre la Tour built the codebase that ended up doing much of the heavy lifting (and he took me to the top of Shasta). Alex Huth, Leila Wehbe, Fatma Deniz, Natalia Bilenko, Robert Gibboni, Michael Oliver, Matteo Visconti di Oleggio Castello, Christine Tseng, and Michele Winter all provided many invaluable feedback during our discussions and lab meetings. And finally, before she left for grad school, Marie Kiesler kept the lab running and fended off all the paperwork that we now have to deal with ourselves.

Thanks would also need to be given to my academic mentors before I came to grad school. I would need to thank David Heeger for taking me, a measly undergrad, into his lab for a summer, and Elizabeth Cutrone, who mentored me there and showed me that grad school was worth it all. Diane Tucker created the SciTech program at UAB, a wonderful community of people that I would never forget. Allan Dobbins introduced me to the concept of neuroscience and studying the brain.

Of course, the real treasure is the friends we made along the way. Together with Alex Levy and Marina Triplett, we commiserated our way through grad school. Along with Alex and Marina, Aaron Stuber also provided much needed nihilism, chaos, and existential angst as we face down the void. My grad school friends James Lucas, Yiqi Cao, Kristen Cotner, and Laura Hallock have been, and continue to be, have been unwavering in their willingness to come on adventures in the mountains. I would also need to thank Anna Zoladz for the only person I know to simultaneously understand and share my passion both for design and space. There

are also friends whom I do not talk to nearly as often as I should: Michael Longmire, Kenneth Gray, and Zac Ingram. They have been an unending source of support, telling me to never stop believing.

Finally, I would of course need to thank my parents, who literally moved halfway across the globe and made sure that I had the brightest future and the best opportunities possible every step of the way.

1 The navigational system of the human brain

1.1. Introduction

Navigation through the world is a critical skill that requires the interaction of many brain systems, including sensory systems for vision and audition, motor systems for movement, and cognitive systems for planning, mapping, and error monitoring (Ekstrom et al., 2014; Garling et al., 1984; Montello, 1998; Spiers and Gilbert, 2015). Navigation is mediated by dozens of distinct functional brain areas that each represent different navigational features (Ekstrom et al., 2014; Garling et al., 1984; Montello, 1998; Spiers and Gilbert, 2015; Spiers and Maguire, 2006, 2007a, 2007b).

Navigation representations in the brain have been studied in parallel in both neurophysiology and neuroimaging. Neurophysiology studies have examined navigational selectivity and representations at the neuronal level. Neurophysiology studies directly record neural activity, either through electrodes or via calcium imaging. These studies are performed mostly in rodents, and, in some infrequent cases, in non-human primates and human epilepsy patients, and have revealed selectivity for many basic navigational features. Parallel to neurophysiology studies, neuroimaging studies have examined navigational representations at the level of voxels. Neuroimaging studies use functional magnetic resonance imaging (fMRI) to capture blood oxygen level-dependent (BOLD) activity (Ogawa et al., 1990), which is hemodynamically coupled to the local neural activity (Heeger and Ress, 2002). These neuroimaging studies are performed in healthy human subjects, and have revealed multiple regions in the human cerebral cortex that are active in scene perception and representation. Here, we summarize the current knowledge from these two fields, and synthesize a tentative integrated view of how navigational features are represented in the human brain. We also identify four unresolved issues in the current understanding of human navigation, and suggest how to close these gaps in understanding.

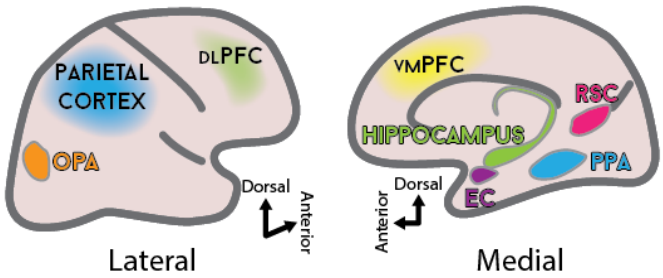


Figure 1.1 The navigational network of the human brain. Neurophysiology and neuroimaging have revealed a network of brain regions that represent navigational information in the human brain. This network includes specialized anatomical and functional regions including hippocampus, entorhinal cortex (EC), retrosplenial cortex (RSC), parahippocampal place area (PPA), and occipital place area (OPA). This network also recruits more generalized regions, such as parts of the parietal cortex, dorsolateral prefrontal cortex (dlPFC), and ventromedial prefrontal cortex (vmPFC).

1.2. The brain decomposes navigation into sub-tasks

Together, neurophysiology and neuroimaging studies have converged on a network of brain regions that represent navigational information (Figure 1.1). This network includes the hippocampus, the subiculum, the entorhinal cortex (EC), retrosplenial cortex (RSC), parahippocampal place area (PPA), occipital place area (OPA), and parts of the parietal and prefrontal cortices. These studies also suggest that the brain breaks down the task of navigation into a set of sub-tasks. These tasks are perceiving local environment, performing sensorimotor transformations, representing an abstract cognitive map of the world, mapping between the perceived environment and the cognitive map, and planning actions and routes (Figure 1.2).

In this review, we make a distinction between "local" and "global" environments. We use "local" to refer to the immediately perceivable environment, and "global" to refer to the world beyond the local environment that cannot be immediately sensed. Representations of the local environment can be constructed from perceptual information alone, while those of the global environment requires an abstract cognitive map. Furthermore, local representations tend to be egocentric, while global representations tend to be allocentric.

1.2.1. Perceiving the environment

Three anterior visual areas – PPA, OPA, and RSC – are key for perceiving and identifying the local environment. Visual perception studies show that PPA is selective for scenes. PPA BOLD activity is higher when subjects view images of places than images of other objects (Epstein et al., 2003; Epstein and Kanwisher, 1998; Park and Chun, 2009). Furthermore, PPA BOLD activity is modulated by both scene structure and the view of the scene, suggesting that PPA likely represents the contents of a scene (Epstein et al., 2003; Park and Chun, 2009). In one experiment, subjects performed two tasks: categorizing rooms (e.g bedroom, kitchen), or identifying navigational affordances (which door is an exit) (Persichetti and Dilks, 2018). The

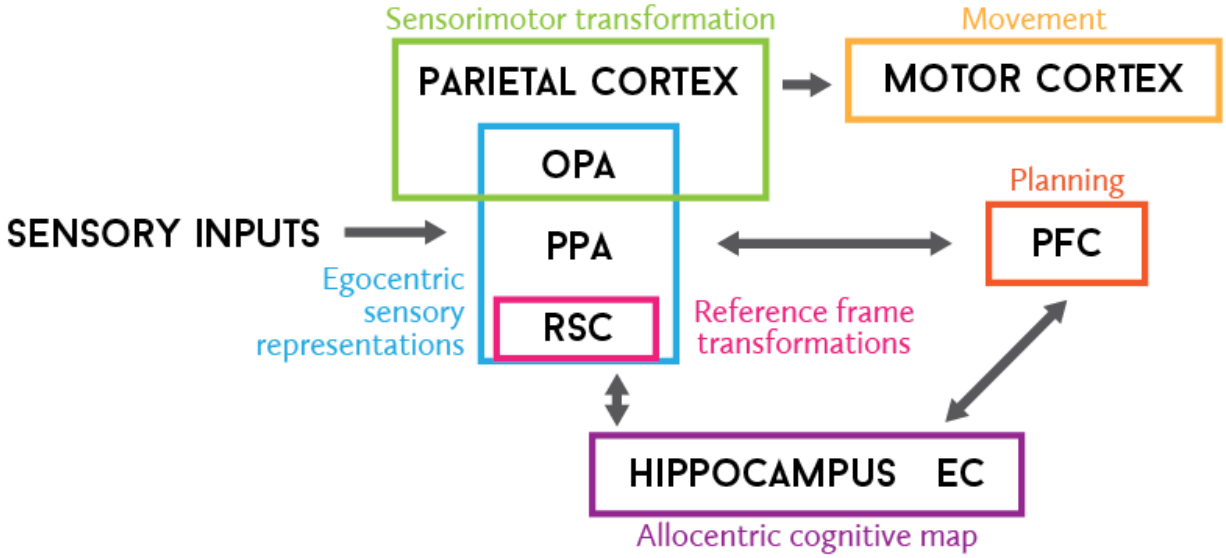


Figure 1.2 A proposed functional organization of the navigation network of the brain. We tentatively divide navigation into a set of sub-tasks, each of which is supported by one or more of the components of the navigational network. The anterior visual areas of OPA, PPA, and RSC are responsible for perceiving and identifying the local environment from visual inputs. They contain egocentric representations of the immediate environment around the subject. The subcortical regions of the hippocampus and EC represent and abstract allocentric map of the world. Their representations may extend beyond the local environment that can be immediately perceived. The RSC, in addition to representing sensory inputs, also performs reference frame transformations to map egocentric sensory-derived representations in anterior visual cortex to abstract allocentric representations in the hippocampus. The prefrontal cortex integrates both sensory egocentric information and abstract allocentric information to make long-range navigational plans and decisions. The OPA and parietal cortex performs sensorimotor transformations to convert sensory representation to actionable features, and the motor cortex plans and executes concrete motor actions to move towards navigational goals.

study found that PPA activity is higher in the room categorization task than in the affordance identification task, suggesting that its role is primarily perceptual rather than sensorimotor. In the same study, OPA displayed the opposite pattern. Nevertheless, OPA actively contributes to scene perception. Indeed, inactivation of OPA with transcranial magnetic stimulation (TMS) disrupts subjects' ability to discriminate between scenes, implying that OPA is casually engaged in scene perception (Dilks et al., 2013). Note that OPA can also be referred to anatomically as the transverse occipital sulcus (Grill-Spector, 2003).

RSC is also implicated in scene perception. In one fMRI study, subjects viewed photos of outdoor scenes in a familiar city (Vass and Epstein, 2013). The identity of locations could be reliably decoded from RSC activity even when photos showed non-overlapping views of the same location. In a similar study, subjects viewed photos of indoor scenes (Kim and Maguire, 2018). In this study, RSC BOLD activity displayed adaptation to disjoint views of the same room, suggesting that these disjoint views elicited similar cortical representations. In another study, subjects viewed snapshots from panoramas of scenes (Park and Chun, 2009). Successive snapshots could be from the same or different panoramas. RSC displayed

adaptation to disjoint snapshots from the same panorama, while PPA did not. These studies suggest that the representation of location in RSC is at least partially view-invariant. On the other hand, the representation in PPA is likely more viewpoint-specific.

The geometry of a scene appears to be represented by both the OPA and RSC. In one study, subjects viewed both intact and fractured photographs of empty rooms (Kamps et al., 2016). Fractured photographs were created by breaking images along environmental edges, and then rearranging the surfaces in pixel space. OPA responded strongly to both intact and fractured images, suggesting that it responds to individual surfaces. On the other hand, RSC activity decreased when subjects viewed fractured images, suggesting that RSC represents higher-level, integrated geometric information about the scene. In another study, subjects viewed rendered videos in which the camera moved through both indoor and outdoor scenes and changed viewpoints (Lescroart and Gallant, 2019). A voxelwise encoding model of the surfaces in the scene is able to predict OPA activity, and the large-scale geometry of scenes can be reconstructed from OPA activity. This result shows that OPA indeed represents surface and large-scale geometry.

The representation of local scene geometry in RSC may be at least partially dissociated from the identity of the scene. In a decoding study, subjects memorized the locations of objects in four geometrically identical buildings in a virtual environment (Marchette et al., 2014). In the scanner, they were asked to imagine facing particular objects. Object identity could be decoded from RSC activity patterns. However, the decoder was more like to confuse objects in geometrically similar locations than those in geometrically dissimilar locations. This confusion suggests that RSC may encode local egocentric scene geometry separately from the identity of the location. Scene geometry likely serves as a cue for scene recognition, as disoriented animals behaviorally confuse geometrically similar corners in an enclosure (Julian et al., 2015). These results suggest that both RSC and OPA may represent purely geometric information about the scene that is detached from scene identity.

RSC also prefers landmarks useful for navigation. In an experiment in which London taxi drivers drove in a virtual replica of London, RSC BOLD activity is elevated when subjects searched for landmarks (Spiers and Maguire, 2006). In another experiment, subjects first learned the layout of a small virtual world, and then navigated through that world in the scanner (Iaria et al., 2007). RSC was active both when subjects were learning the map, and when navigating to particular locations in the world. In navigational learning study, subjects learned the layout of a virtual environment by watching videos that moved through the world (Auger et al., 2015). In this virtual environment, some objects were stationary and were therefore good landmarks, while others randomly moved and were therefore poor landmarks. At the beginning, RSC responded strongly to both good and poor landmarks. As subjects became more familiar with the world, RSC activity remained strong in response to good landmarks, but decreased in response to poor landmarks. This selective change in response suggests that RSC preferentially responds to objects that are useful for navigation. Indeed, in a rodent study in which animals navigated using landmarks, RSC neurons represent vectors anchored only to task-relevant landmarks (Fischer et al., 2020).

Furthermore, lesions to RSC in humans impair navigation using landmarks (Maguire, 2001). Patients with RSC lesions could recognize landmarks and places, but were unable to use landmarks to navigate. This lesion result indicate that RSC may use landmarks to map the current visible environment into the abstract cognitive map.

Together, these anterior visual areas provide a detailed representation of the immediate visible world. While all three areas contribute to scene perception, each has a preferred function. PPA identifies the scene and forms a low-level visual representation of the scene, and OPA builds geometrical information about the scene. Both PPA and OPA function in egocentric space, and are mostly driven by sensory inputs. They function on immediate local scales, representing features in the visible environment. Hence their activities are modulated by viewpoint within scenes and elements in a scene. RSC then uses this information to construct a more perceptually-invariant representation of the scene.

1.2.2. Sensorimotor transformations

Sensorimotor transformations are likely performed by OPA and also parts of the parietal cortex. These transformations convert perceptual information to representations that are useful for moving through or interacting with the local environment. (Dilks et al., 2013) disrupted OPA activity with TMS in a perception task, and found that doing so degraded recognition of scenes but not objects, demonstrating that OPA is casually involved in scene perception. (Julian et al., 2016) again disrupted OPA activity with TMS, this time in a virtual reality navigation task. TMS of OPA impaired the accuracy of navigation to locations anchored to environmental boundaries. These TMS studies causally implicate OPA in representing information needed to actively navigate in the local environment.

In addition to these TMS studies, a fMRI study found that OPA BOLD activity is higher when subjects searched for possible exits from a room than when they simply categorized the room (Persichetti and Dilks, 2018). This difference indicates that the need for active navigation engages activity in OPA beyond perceptual demands. In another fMRI study, subjects viewed both rendered images and photographs (Bonner and Epstein, 2017). OPA activity was correlated with navigational affordances in the scenes, and also provided the most accurate reconstruction of affordances when compared with PPA and RSC. Taken together, these results suggest that OPA may not only represent perceptual information, but also likely information needed for actively navigating through the scene.

Outside the OPA, in the parietal cortex, the superior parietal lobule (SPL) also appears to represent affordance information. In a high-field fMRI study, subjects viewed mazes that had exits places in one of 8 possible orientations (Gourtzelidis et al., 2005). SPL appeared to represent the direction of the exit when subjects solved the maze in the scanner. In animal studies, many posterior parietal cells represent conjunctions of multiple features. These conjunctions include movements in particular locations (Mcnaughton et al., 1989; Sato et al., 2006) and goal direction relative to head direction (Wilber et al., 2014). These results suggest that OPA and the posterior parietal cortex are important for performing sensorimotor transformations by mapping perceptual information into actionable reference frames.

1.2.3. Representing an abstract cognitive map of the world

RSC, OPA, and PPA are anterior visual areas whose representations are perceptually driven and largely reflect only the local environment. These representations are complemented by a cognitive map represented in the hippocampal formation. This cognitive map can represent both the local and global environments, including objects and places that cannot be immediately perceived.

1.2.3.1. Own location

To make use of a map, an animal must be able to locate itself on it. The rodent hippocampus contains place cells that represent the animal's position in the local environment by firing only when the animal is in a particular location (O'Keefe and Dostrovsky, 1971; O'Keefe and Nadal, 1978). Place cell firing fields exist at multiple scales to give varying degrees of accuracy (Kjelstrup et al., 2008). These firing fields are remapped between different local environments, and so the same physical circuitry can be used to represent position in multiple environments (Muller and Kubie, 1987). In humans, electrophysiology in epileptic patients have also revealed the existence of place cells (Ekstrom et al., 2003). Because non-invasive modalities currently lack single-cell resolution, human neuroimaging studies cannot directly observe place cells. Nevertheless, numerous fMRI studies have demonstrated that the human hippocampus is activated during spatial memory tasks and plays an important role in human navigation (Burgess et al., 2002; Epstein et al., 2017; Ghaem et al., 1997; Maguire et al., 2006; Parslow et al., 2004; Rosenbaum et al., 2004).

Along with position, an animal must be able to represent its orientation to make effective use of a map. The hippocampal position code is complemented by an allocentric direction code in multiple brain regions. Rodent studies have found head direction (HD) cells that fire whenever the animal faces a particular allocentric direction. These HD cells are found in the postsubiculum (Taube et al., 1990), RSC (Chen, Lin, Barnes, et al., 1994; Chen, Lin, Green, et al., 1994), posterior parietal cortex (Chen, Lin, Barnes, et al., 1994; Chen, Lin, Green, et al., 1994), thalamus (Mizumori and Williams, 1993; Taube, 1995), and striatum (Wiener, 1993). Both HD (Taube et al., 1990) and place cells (Muller and Kubie, 1987; O'Keefe and Conway, 1978; O'Keefe and Speakman, 1987) are environmentally-anchored: their firing fields move if distal cue cards are moved, and they show similar representations in visually identical rooms.

Note that RSC HD cells can be locally or globally anchored (Jacob et al., 2017). In this experiment, rats were moved between two identical rooms connected by an opening. Locally anchored HD cells followed a cue on the room wall, while the globally anchored cells did not. Thus, it is likely that HD cells may be anchored to reference frames at varying spatial scales.

In humans, fMRI provides partial support for a representation of allocentric head direction. In one fMRI experiment, subjects viewed outdoor scenes taken in a familiar city (Vass and Epstein, 2013). The allocentric facing direction (e.g. north, west) could be decoded from BOLD activity in the presubiculum, suggesting the presubiculum in humans may also contain a representation of global direction.

1.2.3.2. Other aspects of the cognitive map

Place cells and head direction cells together represent the position and orientation of the animal on the cognitive map. In order to be useful, this map must also include representations of the environment, such as objects and landmarks, boundaries, and other animals. Indeed, the rat hippocampus contains landmark vector cells whose firing fields are anchored to a specific object instead of an absolute location in the environment (Deshmukh and Knierim, 2013). These landmark vector cells fire when an object is at a particular location and direction from the animal. EC also contains object vector cells whose firing fields are anchored objects in the environment (Deshmukh et al., 2012; Deshmukh and Knierim, 2011; Savelli et al., 2008, 2017). These object vector cells in the EC use both allocentric (Høydal et al., 2019) and egocentric reference frames (Wang et al., 2018). Furthermore, unlike hippocampal landmark vector cells, object vector cells in EC generalize to objects of different shapes and sizes (Høydal et al., 2019).

EC also contains boundary vector cells that represent environmental boundaries (Solstad et al., 2008). These cells fire when the animal is next objects that impede movement, and their spatial firing fields scale with size of the barrier, and disappear if the barrier is taken away. Similar boundary representations are also found in the subiculum (Lever et al., 2009).

In addition to environmental features, this cognitive map may also include other agents. In two experiments, bats (Omer et al., 2018) and rats (Danjo et al., 2018) observed other bats or rats performing tasks in a familiar enclosure. Both experiments found that hippocampal CA1 cells display spatial firing fields for other animals. Furthermore, some of these cells were also canonical place cells, i.e., they also display a spatial firing field for the animal's own position. Note that in these cells, the spatial firing fields for the animal itself and for other animals may not be spatially co-localized. Nevertheless, the same neural circuitry is used to represent both the animals' own location and the location of other agents, suggesting that they are represented on the same cognitive map.

1.2.4. Measuring distances on the cognitive map

The animal must be able to manipulate and interpret its cognitive map to make use of it. A key basic manipulation is to measure distances on a map. EC provides a metric for measuring space on the cognitive map in an allocentric reference frame. Grid cells in EC fire at regular spatial intervals on a hexagonal lattice (Brun et al., 2008; Fyhn et al., 2004; Hafting et al., 2005). Grid cells have different spatial phases and periods, and their lattice points tile space. By counting the number of firing fields traversed, an animal would be able to estimate distance. Furthermore, because grid cells have different phases and periods, they can be mathematically combined to derive place cell fields (Solstad et al., 2006), providing a method for calculating position. The activities of some grid cells are be modulated by the speed and head direction of the animal (Sargolini et al., 2006), suggesting that they represent a conjunction of distance and other spatial information.

Grid cells are also observed in humans using electrophysiology (Jacobs et al., 2013). Neuroimaging has also produced indirect evidence for grid-like representations of space (Bellmund et al., 2016; Doeller et al., 2010), however these results are tenuous and difficult to replicate. These EC representations of distances may also be supplemented by hippocampal representations. In a study in human epilepsy patients, hippocampal theta power was found to scale with the distance traveled (Bush et al., 2017).

The metrics provided by grid cells are not necessarily environmentally-anchored or need a complete cognitive map. Grid cell firing fields persist in darkness in the absence of sensory inputs (Hafting et al., 2005). Furthermore, entorhinal grid representations arise quickly when animals are introduced to new environments (Hafting et al., 2005; Høydal et al., 2019). On the other hand, hippocampal cells only develop firing fields after repeated exposure to an environment (Høydal et al., 2019). Thus, EC likely provides a set of sensory-derived representations to construct and use the abstract allocentric map.

1.2.5. Transformations between reference frames

The cognitive map in the subcortical structures is in an allocentric reference frame anchored to the environment. Sensory inputs in the cortex, on the other hand, are in egocentric reference frames referenced to the subject. For example, visual inputs are in retinotopic coordinates centered on the fixation, and arm positions are referenced to the body. Transformations between reference frames are needed to map these sensory representations on to the abstract cognitive map in subcortical structures.

Reference frame transformations are likely mediated by the RSC. RSC contains both egocentric and allocentric representations, and many cells represent the conjunction of multiple features. In one study, rats ran inside a track contained within a larger visible room (Alexander and Nitz, 2015). RSC neurons represented the conjunction of the animal's position within a track, i.e., a local reference frame, and the animal's position within the room, i.e., a global reference frame. In humans, RSC BOLD activity shows adaptation to global head direction when subjects viewed snapshots of a familiar virtual environment (Shine et al., 2016). This result suggests that RSC may contain allocentric representations of head direction in addition to the egocentric representations of the local scene. The conjunctive representations found in rat RSC, and the possible presence of both allocentric and egocentric representations (see section 1.2.1), suggest RSC may perform reference frame transformations (Alexander and Nitz, 2015; Nitz, 2012; Whitlock et al., 2008) and scene integration (Epstein, 2008; Vann et al., 2009). This mapping role is supported by RSC's strong anatomic projections to the hippocampal formation in both rodents (Wyss and Van Groen, 1992) and non-human primates (Kobayashi and Amaral, 2003, 2007).

1.2.6. Planning actions and routes

Planning actions and routes is largely performed by the prefrontal cortex (PFC). The PFC integrates both the local environment from sensory inputs and the allocentric cognitive map beyond the current visible environment to make long-term, goal-oriented plans.

the Navigational System of the Brain

Several regions of prefrontal cortex are engaged in navigational planning and representing goals in rodents (Rossi et al., 2012; Yoon et al., 2008). (Granon and Poucet, 1995) found that rats with ventromedial prefrontal lesions were unable to adapt to new starting locations in a Morris water maze. In one experiment, rats must navigate to an unmarked zone in an enclosure to receive a food reward (Hok et al., 2005). In this task, medial PFC (mPFC) neurons in the prelimbic and infralimbic areas were found to exhibit spatial firing fields. These spatial firing fields are disproportionately clustered around the goal, suggesting that mPFC encodes the goal for path planning. In another rat study, rats must decide the direction of a reward given odor cues (Feierstein et al., 2006). Over half of OFC neurons recorded were found to represent a choice of direction or direction of the goal. Furthermore, a subset of these OFC neurons encodes the conjunction of the chosen location and the task outcome, suggesting that OFC may be updating choice parameters for future trials.

Multiple human neuroimaging studies find that PFC is engaged in navigation tasks. In one experiment, subjects drove in a video game version of London while BOLD activity was recorded, and described their thought processes after scanning (Spiers and Maguire, 2006, 2007b). During the periods in which subjects were thinking about navigational goals, ventromedial and ventrolateral PFC activity was significantly elevated. In another experiment, subjects memorized locations on a virtual track, and navigated to cued locations in the scanner (Brown et al., 2016). Here, the goal direction can be decoded from frontopolar cortical activity patterns, and also from hippocampus. In additional to representing foals, PFC is likely engaged in manipulating and searching the cognitive map. In one experiment that sought to examine path planning, subjects watched videos of routes through Soho, London, and were asked to make turn decisions at key junctions (Javadi et al., 2017). Activity in inferior lateral prefrontal cortex increased with the complexity of a breadth-first search over a road graph. In a similar task, subjects viewed mazes and were asked to make a path choice (Kaplan et al., 2017). Rostrodorsal medial prefrontal cortex, superior frontal gyrus, and lateral frontopolar cortex were all modulated by the difficulty of the choice.

The importance of PFC to navigation is also highlighted by a case study of a patient with a ventromedial PFC lesion (Ciaramelli, 2008). The patient had intact knowledge of landmarks and routes in a familiar town. However, when tasked with navigating somewhere, the patient frequent forgot where he was going, and gravitated to a few familiar locations. Performance improved when the subject was reminded of the goal, suggesting that successful navigation requires vmPFC to maintain goal information in working memory.

Multiple regions outside of the PFC are engaged in tracking progression towards a goal. In one fMRI study, subjects were first shown a goal location in an empty virtual world. Then, the goal marker is hidden, and subjects were asked to navigate to the goal location (Sherrill et al., 2013). RSC is active both when subjects were shown the goal, and also when subjects navigated towards the goal, indicating that it likely encodes information relevant to navigation towards a memorized goal. In another fMRI study, subjects watched multiple computer-generated videos in which the camera moved through a virtual forest (Chrastil et al., 2015). Subjects then judged whether paths in two successive videos arrived at the same

point. RSC, hippocampus, and the parahippocampal cortex were all active when subjects were successful in judging endpoints, suggesting that these areas may be responsible for representing a homing vector.

These results are also reflected in electrophysiology. In a series of rodent experiments, rats ran along a track, and neural activity from posterior parietal cortex (PPC) and RSC were recorded (Alexander and Nitz, 2017; Nitz, 2006). Cells in PPC were selective for particular locations along a track (Nitz, 2006, 2009), while cells in RSC displayed spatially periodic firing patterns along a track (Alexander and Nitz, 2017). In particular, the RSC cells displayed the same spatial periodicity across differently shaped tracks, indicating that they are not environmentally-anchored, but rather reflect some abstract representation of a path. These results in both humans and rodents suggest that the roles of these regions are not merely perceptual, but are involved in representing top-down navigational planning information.

Subcortical structures are also likely engaged in representing goal-related information. Distance and bearing to a navigational goal modulates activity in both the hippocampus (Howard et al., 2014; Sherrill et al., 2013; Spiers and Maguire, 2007b) and EC (Brown et al., 2016; Chadwick et al., 2015). When subjects navigated along a virtual track, goal identities, along with locations along the planned path, are also decodable from hippocampal activity (Brown et al., 2016). In the same study that suggested RSC and parahippocampal cortex may represent a homing vector, hippocampal activity was also found to track a homing vector (Chrastil et al., 2015). Furthermore, in rats perform a goal-directed navigation task, grid cell firing fields were found to be modulated by goal direction (Sarel et al., 2017), indicating that goal direction is represented in EC neurons.

1.3. Task effects on navigational representation

Navigation is an inherently dynamic task composed of multiple subtasks, and requires the navigator to respond to changing environment. In addition to revealing multiple navigation areas and representations, neurophysiology studies have also revealed significant task effects on navigational representations.

The effects of active navigation on representation are well known in rodent electrophysiology studies. HD cells in the postsubiculum demonstrate decreased firing rates when rats are moved around by hand rather than when they were allowed to freely roam (Taube et al., 1990). Hippocampal place cell activity is heavily dependent on active locomotion. In one study, rats were first allowed to actively run along a circular track (Song et al., 2005). Then, they were passively moved along the track by a motorized cart. Place cell firing fields were remapped when rats were passively moved, despite the identical environment. Then, rats were allowed again to actively run along the track, and places cells restored their original firing fields. This remapping suggests that the same environment may have vastly different representations depending on whether an animal is allowed to navigate. In another rodent study, virtual reality (VR) was used to dissociate visual inputs from locomotion (Chen et al., 2013). First, spatial firing fields were established for places cells while rats actively ran in

an VR environment with landmarks. Then, rats were presented with the same landmarks when they were not moving. Only 25% of cells showed the same spatial firing fields in this condition. Indeed, in most rodent studies, data from periods during which rats are not actively moving are discarded, reflecting an implicit assumption that navigational representations require active locomotion.

Similar effects are also seen in RSC. Navigational representations are shown to be nonlinear with respect to multiple experimental factors in rodents. In a VR-based task, rats used landmarks to determine the position of a reward (Fischer et al., 2020). During this task, RSC neurons represented vectors anchored to landmarks. In a control condition, rats were freely running but not performing this task. Neurons that are landmark-selective during the task did not show landmark selectivity when the same landmarks were presented outside of the task context, regardless of whether the rats were moving or not. Furthermore, the population code for the rat's position is markedly sharper during the active navigation task. Visual-motor integration of RSC activity amplitude is also found to be nonlinear. RSC activity in when rats saw landmarks while running is higher than the sum of activity when rats were only running and when they saw landmarks while not running. An active task further increases RSC activity above levels from when the rat is freely running and saw landmarks. These results demonstrate that RSC representations change dynamically with tasks, and representations during passive viewing may not reflect representations during active navigation.

Another virtual reality navigation experiment demonstrated that representations across the rat dorsal cortex are task-dependent (Pinto et al., 2019). In this experiment, rats ran down a T-maze and performed a visually-guided task, a memory-guided task, and an accumulating evidence task to acquire a reward. In the visually-guided task, rats simply followed a visual cue to the reward. In the memory-guided task, the cue was presented to the animal only at the beginning of the trial, and animals must remember the cue location. In the accumulating evidence task, mice must count cues appearing on both sides of the track, and go to the side with more cues. In the more demanding tasks, the activity across the dorsal cortex were more decorrelated than in less demanding tasks. Because the visual environment and motor behavior were similar across tasks, this correlation change can only be attributed to the tasks eliciting different computational demands. These results suggest that eliciting ecologically valid navigational representations require not only the complete visual-motor feedback loop, but also an active navigation task.

1.4. Unresolved issues in understanding navigation in the human brain

We have developed a global view of the brain's navigational network using both the neuroimaging and neurophysiology literature from humans and rodents. However, there exist several significant gaps in our current understanding of navigational representations in the human brain. These are the representation of space, a gap between neurophysiology and

neuroimaging results, the ecological validity of most neuroimaging navigation experiments, and the inability of common neuroimaging analysis methods to capture high-dimensional representations in complex experiment.

1.4.1. The representation of space

The navigational network of the brain includes multiple representations of space. Both hippocampal place cells and entorhinal grid cells provide strong evidence for a metric representation of space. In a metric representation, the world is represented on a Euclidean map in which all spatial structure is preserved, and objects can be assigned coordinates. However, there has been some evidence in favor of a topologic representation on larger scales. In a topologic representation, the relationship between locations, but not necessarily their exact spatial structure, are represented on a graph, and objects are represented relative to each other instead of given positions in a coordinate frame.

Human behavioral studies in VR mazes suggest that humans learn a topological representation for the global environment, with locally metric information (Chrastil and Warren, 2014). In one study, subjected learned the layout of a non-Euclidean maze in which two distal parts were joined by a wormhole (Warren et al., 2017). Subjects did not readily notice the non-Euclidean geometry of the maze. When asked to draw the shape of the maze, subjects accounted for these wormholes by distorting the map. This behavioral blindness to non-Euclidean geometry suggests that a metric Euclidean representation may not be a core component of the cognitive map. Indeed, neuroimaging suggests that activity in the right hippocampus tracks topologic measures about road segments in road networks (Javadi et al., 2017). Furthermore, electrophysiology in rats show that hippocampal place cells preserved the topologic order of their place fields in a linear track that changed physical shape while preserving topology (Dabaghian et al., 2014). These results suggest that the navigation system of the brain may contain topologic representations that are distinct from the metric representations used by canonical place and grid cells.

We argue that a metric representation is only tractable up to a certain scale. It is improbable that the brain can maintain a full metric representation of the entire world. Furthermore, a metric representation is arguably only necessary for the immediate visible space. It is likely that there is a gradient of representations, or multiple co-existing representations. Here we make a distinction between “simple” and “complex” environments. A simple environment is largely visually contiguous, and can be navigated perceptually, such a single room. A complex environment is visually disjoint, and require mental representations of locations that cannot be perceived, such as an entire building. Simple environments may be represented more metrically with their own reference frames, while complex environments may be represented more topologically. Complex environments can be composed from multiple simple environments to form a hierarchy of representation. In this hierarchy, simple environments are metrically represented, and different metric representations are linked by a topological representation (Chrastil and Warren, 2014). This possible difference in representation might

explain why hippocampal place cells exhibit remapping between distinct environments. Place cells may form the circuitry for local metric representations, and are re-used in different metric reference frames associated with different simple environments.

Note that most studies have overwhelmingly only examined metric representations in terms of place and grid cells. This bias may be due to experimental constraints. Neurophysiology and neuroimaging studies generally use experiments in which there are only simple environments, such as a single enclosure for rodents, or viewing single images for humans. Thus, they are not be able to examine the representation of complex environments that are composed of multiple visually disjoint simple environments. There is therefore a need to better examine the representation of space in complex environments, and investigate the possible different modes of spatial representation.

1.4.2. Linking neurophysiology and neuroimaging

While both human neuroimaging studies and non-human electrophysiology studies aim to study how the brain represents navigational information, few neuroimaging studies have directly demonstrated the same results about representation as neurophysiology studies. Neuroimaging studies typically do not seek to answer the same questions as neurophysiology studies. Furthermore, they typically use different stimuli and tasks from neurophysiological studies, and manipulate different experimental variables, and thus examine different representations. To our knowledge, only entorhinal representations found by neurophysiology have been tentatively replicated by neuroimaging studies (Doeller et al., 2010; Nau et al., 2018). Even these replications are indirect, and rely on many assumptions about the properties of grid cells.

The different results from neurophysiology and neuroimaging are not necessarily contradictory, and indeed can be seen as complementary. Nevertheless, it is important to keep in mind that the vast majority of electrophysiology studies are performed in rodents, and their results cannot be assumed to translate completely to humans. Organizationally, the rodent neocortex differs significantly from that of primates. For example, areas OPA and PPA are not identified in rodents and have no clear homologies. The lack of clear homologs is not surprising, given the vast differences between human and rodent brains.

Representationally, rodents may adopt different strategies than primates. A recent preprint in freely-moving macaques shows that the primate hippocampal formation represents the conjunctions of many features (Mao et al., 2020). These primate representations are dominated by allocentric facing direction and tilt selectivity, while place and grid cells constitute a much lower proportion of the population than in rodents. This difference may reflect the dominance of vision in primates, and suggests that primate representations contain different biases than rodent representations.

Nonetheless, navigation is a basic and crucial skill for all mobile animals. It is likely that many core representations and functions are conserved across species. However, cortical navigation areas may have become more specialized in primates, which may also use more

complex navigational representations and strategies than rodents. Therefore, there is a need to reconcile the results between neurophysiology and neuroimaging for a more complete understanding of navigational representation in the human brain.

1.4.3. Ecological validity of neuroimaging navigation stimuli and tasks

The brain is a nonlinear dynamical system that may operate differently under simplified conditions than under natural conditions (MC-K Wu et al., 2006). Simplified and controlled experiments often recruit fewer brain areas, or recruit them differently than in daily life (Cremers et al., 2017; Desmond and Glover, 2002; Hasson et al., 2010; Matusz et al., 2019; Mumford and Nichols, 2008; MC-K Wu et al., 2006). Thus, in order for results to generalize to the real world, experiments must be ecologically valid. Yet, most human neuroimaging navigation studies have used stimuli and tasks with poor ecological validity.

1.4.3.1. Stimuli

In choosing stimuli, neuroimaging experiments typically try to remove, clamp, or randomize variables that are not directly related to the hypothesis of interest. Doing so reduces the effects of stimulus correlations, in which confounding variables may bias results. The result is tightly controlled and artificial stimuli that most likely do not reflect real-world conditions. Examples include barren environments (Sherrill et al., 2013), fractal images on a circular track (Brown et al., 2016), and computer-rendered rooms with arbitrary paths on the floor (Persichetti and Dilks, 2018).

However, correlations between different features are a fundamental aspect of the natural world (Field, 1987). For example, curvature and animacy are strongly correlated in the natural world. Animal silhouettes tend to be curved, while the silhouettes of inanimate objects tend to possess more straight lines and sharp corners. Removing correlations is an imperfect solution, because some features are correlated in a way that cannot be removed, and other feature correlations are unknown (Lescroart et al., 2015). Thus, the artificial stimuli used in many neuroimaging studies may not elicit the same activities as more naturalistic stimuli.

Neuroimaging navigation studies have been steadily improving the ecological validity of stimuli. In more recent experiments, there has been a trend towards more naturalistic stimuli, such as photographs (Vass and Epstein, 2013) and videos (Howard et al., 2014) from real-world locations.

1.4.3.2. Tasks

Despite the improvement in the ecological validity of stimuli, the tasks used in the vast majority of neuroimaging navigation studies remain highly controlled and artificial. Many tasks do not let subjects interact with the stimuli. Subjects view predetermined photos or videos, and perform a passive navigation-related task, such as judge direction (Vass and Epstein, 2013), make a navigational decision (Howard et al., 2014), track distance traveled (Chrastil et al., 2015), or imagine being somewhere (Marchette et al., 2014). Sometimes the task is even unrelated to navigation, such as color comparison or object category

detection (Bonner and Epstein, 2017). These tasks are overwhelmingly singularly focused on perception, memory, or decision-making. This singular focus may not reflect the demands of active, dynamic navigation through an environment.

Navigation is an inherently active task in which the subject's actions affects what the subject perceives as the subject navigates through the environment. Active navigation engages the full sensory-motor-planning loop, while passive tasks can only examine a subset of the brain systems needed for naturalistic navigation. Active navigation dynamically changes the local environment around the subject. It also dynamically changes the state of the navigational task, such as progression towards a goal. These dynamics cannot be captured by fixed stimuli and passive tasks. Indeed, rodent studies have shown that navigational representations are elicited only when rats are actively navigating even when stimuli are held constant (see section 1.3).

Very few neuroimaging studies use active navigation tasks either due to the difficulty of implementing such a paradigm in MRI, or of analyzing such data. Of the studies that did use active navigation, many simplify the environment to be sparse and unnatural (Brown et al., 2016; Doeller et al., 2010; Sherrill et al., 2013). We know of only one study that used an active navigation task in a naturalistic environment that is also fully interactive (Spiers and Maguire, 2006).

Neurophysiology studies have long used active navigation paradigms. Rats freely roam in enclosures (Fyhn et al., 2004; O'Keefe and Conway, 1978; Sargolini et al., 2006; Taube et al., 1990), tracks (Alexander and Nitz, 2015) or mazes (Nitz, 2006), and more recently in virtual reality (Aronov and Tank, 2014; Harvey et al., 2009; Hölscher et al., 2005). Active navigation in virtual environments have also been used in both human (Bush et al., 2017; Ekstrom et al., 2003; Jacobs et al., 2013) and non-human primate neurophysiology studies (Sato et al., 2006).

Rodent neurophysiology studies have demonstrated that navigational representations are heavily task-dependent (see section 1.3). These rodent results suggest that passive tasks used in the majority of neuroimaging studies are unlikely to elicit the same activity as active navigation, biasing the conclusions. The ecological validity of neuroimaging experiments using controlled stimuli and passive tasks are therefore highly questionable. There is a need for more ecologically valid experiments that include not only naturalistic stimuli, but also active navigation tasks that emulate real-world navigation.

1.4.4. Limitations in current data analysis methods

The dearth of active navigational tasks in human neuroimaging studies is closely related another issue. Namely, this is the limited power of common analysis methods, such as contrasts, adaptation, statistical parametric mapping (SPM) (Friston, 1994b), multivariate pattern analysis (MVPA) (Norman et al., 2006), and representational similarity analysis (RSA) (Kriegeskorte et al., 2008). The methods are not powerful enough to reveal high-dimensional representations in cortical activity, particularly under naturalistic conditions and in a dynamic environment. Contrasts and SPM only capture the existence of activity, and not tuning or representation, and thus cannot reveal much about cortical representations or computations. While adaptation may be used to infer tuning to individual features, it cannot

efficiently map the representation of many features across the brain. Multivariate analyses such as MVPA and RSA only reveal the existence of representation in a cortical region, and perhaps the type of representation in the case of RSA, but still cannot fully describe these representations. These multivariate analysis methods also require differential tuning across voxels, and cannot make conclusions about individual voxel tunings or representation. Thus, MVPA and RSA inherently discard spatial information. Furthermore, they cannot disambiguate stimulus correlations from true representations. Finally, these methods discard temporal information before analysis, and thus are ill-suited for dynamic tasks such as driving. These methods are also designed to test individual hypotheses, and cannot fully capture and model the representation of multiple types of navigational information across the brain in a single experiment, and at the finest resolution available in MRI: the voxel.

Human neuroimaging navigation experiments are therefore typically designed to maximize statistical power with respect to a particular hypothesis of interest. Experiments also focus almost exclusively on functional regions-of-interest (ROIs) to further maximize statistical power. ROI-based approaches discard potentially useful information represented at single voxel level. These approaches cannot easily identify functional subdivisions within ROIs, and are likely to miss ROIs that represent novel stimulus- or task-related variables, or variables that are inconsistent with prior expectations. These analyses rarely produce high-resolution quantitative models that encompass the cortex and describe the representation of many different types of information in naturalistic conditions. Crucially, these analyses are unable to predict brain activity and generalize to novel data. Generalization is critical to testing the ecological validity of conclusions by showing that the results are applicable beyond the dataset from which they are drawn.

These analytical and experiment design limitations make it difficult to test a large number of hypotheses at the same time across the brain. They have led to the widespread use of non-ecologically valid passive tasks that may skew conclusions about navigational representations, and also many different claims about navigational representations in different brain areas. It is hard to determine whether competing claims are mutually exclusive or can co-exist, or the relative importance of co-existing hypotheses. Many of the representations discovered from neurophysiology in rodents have thus also not been examined in neuroimaging. Therefore, there is a need for better analysis methods that can accommodate the complex data from active navigation experiments, and to efficiently test many hypotheses about representations in the same setting to determine their validity.

1.5. Overcoming these limitations

Neuroimaging studies of navigation requires a new paradigm to overcome these limitations and answer the unresolved issues. The paradigm needs an experiment with ecologically valid stimuli that simulates real-world conditions at both local and global scales, and active navigation tasks that are not merely perceptual or decision making to fully engage the navigational system of the brain. Such an experiment would not only be more ecologically valid, but also be closer to tasks used in rodent studies, enabling us to better test hypotheses derived from the rodent navigation literature.

There is also a need for a more powerful analysis method that can simultaneously test many hypotheses about representations in data from complex tasks, and identify the relative importance of each hypothesis. There is need for the method to capture what features are represented by brain activity, and also the cortical tunings for these features at the highest possible spatial resolution offered by fMRI. The analysis method must account for correlations during data analysis, instead of stimulus design, to identify which specific features elicit brain activity. Finally, this method must produce robust quantitative computational models that accurate predict brain activity and generalize to new data. A new paradigm will enable us to better bridge the neurophysiological and neuroimaging literature on navigation, to better understand how the human brain represents information to navigate through the natural world.

A better understanding of navigation will also have significant clinical applications in addition to scientific knowledge. Navigational deficits due to neurodegenerative diseases severely impact quality of life. Behavioral studies suggest that navigation dysfunction might be a sensitive biomarker for aging and for neurodegenerative diseases (Gazova et al., 2012; Laczó et al., 2018; Lester et al., 2017; Vlcek, 2011). Thus, a better understanding of how navigation is represented in the brain has both scientific and clinical impacts.

2 Building a dynamic and interactive virtual world for naturalistic navigation in fMRI

2.1. Introduction

Navigation is a crucial skill that elicits rich activity across many functional regions and networks across the brain. Neurophysiology studies have revealed many of the brain's foundational representations of navigational information, such as places cells (O'Keefe and Dostrovsky, 1971), grid cells (Fyhn et al., 2004; Sargolini et al., 2006), and head direction cells (Taube et al., 1990). However, animals are only imperfect models of human navigation, and human studies are critical to understanding how the human brain supports navigation.

Functional magnetic resonance imaging (fMRI) produces rich data of the hemodynamic activity across the brain, and has enabled the non-invasive study of brain activity in healthy human subjects (Ogawa et al., 1990). A rich body of literature has revealed a complex network of both cortical and subcortical areas that support navigation. These areas include the parahippocampal place area (Epstein et al., 1999), occipital place area (Dilks et al., 2013), retrosplenial cortex (Aguirre et al., 1998), parts of the prefrontal cortex (Brown et al., 2016), the hippocampus (Burgess et al., 2002; Maguire et al., 2006; Parslow et al., 2004), and entorhinal cortex (Doeller et al., 2010).

However, the current understanding of navigational representation in the human brain is limited by two intertwined barriers. The first barrier is a data analysis barrier. Traditional analysis methods, such as statistical parametric mapping (Friston, 1994b), multivariate pattern analysis (Norman et al., 2006), and representational similarity analysis (Kriegeskorte

et al., 2008), discard significant amounts of spatial information in fMRI data. Thus, they cannot construct detailed functional maps of representation in the brain. Furthermore, these methods are ill-suited for time-varying analyzing data collected from dynamic tasks.

The second barrier is an experimental design barrier. Experiments are typically designed to maximize signal with respect to individual hypotheses. This leads to tightly constrained and artificial experiments with highly unnatural stimuli. The data analysis methods greatly constrain the complexity of experiments, and the tightly controlled experiments do not reflect real-world cognitive demands found in everyday life (Cremers et al., 2017; Desmond and Glover, 2002; Hasson et al., 2010; Matusz et al., 2019; Mumford and Nichols, 2008; MC-K Wu et al., 2006).

Furthermore, in the vast majority of experiments, the subjects cannot interact with the stimuli. Interaction with the stimuli is a crucial component of navigation. Navigating through the world necessarily changes what we perceive. In most experiments, however, subjects cannot do so (See section 1.4.3.2). Instead, they perform some passive perceptual, decision-making, or even unrelated task while they watch fixed stimuli.

This lack of active navigation tasks also reflects the technical difficulty of implementing an active navigation paradigm. Standard stimulus presentation libraries for fMRI and psychophysics only have rudimentary graphics capabilities and cannot offer interactivity with the subject beyond simple button presses. Mainstream stimulus presentation libraries such as PsychoPy for Python (Peirce, 2007) and the Psychophysics Toolbox for Matlab (Brainard, 1997) are rudimentary in nature, focusing on the presentation of text and images. These scripting libraries cannot emulate a naturalistic virtual world and allow subjects to navigate in them.

In a few active navigation experiments, custom virtual worlds (Brown et al., 2016; Sherrill et al., 2013), along with rendering and experiment logic, are created from scratch, an exceedingly time-consuming task. These bespoke worlds are nonetheless sparse and tightly controlled. The brain is a highly nonlinear system, and brain activity during simplified tasks may not reflect brain activity during active, naturalistic navigation. Indeed, rat studies show that cortical neurons in RSC and parietal cortex only show navigational activity when the rat is actively performing a navigation task (Fischer et al., 2020).

To overcome the first barrier, our lab had developed voxelwise modeling (VM), a systems identification method adapted from neurophysiology (Naselaris et al., 2011). VM does not discard spatial information, and can analyze brain activity in response to complex naturalistic stimuli. This efficacy of this method has been demonstrated in multiple high-impact publications in both visual and language neuroscience experiments (Çukur et al., 2013; Deniz et al., 2019; Huth et al., 2012, 2016; Lescroart and Gallant, 2019; Nishimoto et al., 2011). In doing so, VM also overcomes the experimental design limitation, and does not constrain experiments to tightly controlled, artificial tasks. Here we present an experimental framework that overcomes the second barrier, and enables us to study active navigation in a naturalistic world with fMRI.

2.2. A naturalistic active navigation framework

In this novel experimental framework, we leverage Unreal Engine 4 (UE4, Epic Games, Cary, NC), a widely-used, state-of-the-art game engine, to create the most naturalistic active navigation paradigm to date. The flexibility, ease-of-programming, and wide availability of resources of UE4 makes it uniquely suited for interactive, naturalistic experimental paradigms such as active navigation. UE4 offers an extensive library and ecosystem for creating interactive content. The C++ development environment provide an easy-to-program and highly efficient framework for implementing experimental logic. The Blueprint visual scripting language provides an intuitive method for automating level design and creating content. The Unreal Motion Graphics UI designer provides an intuitive visual designer for user interface design.

In addition to the core Unreal Engine, we also build heavily upon Carla, a UE4 plug-in developed to generate simulated data for training autonomous vehicle controllers, to facilitate some aspects of the experiment (Dosovitskiy et al., 2017). Furthermore, we make use of many off-the-shelf 3D assets to construct the virtual world.

In this paradigm, subjects drive in a large virtual city using an MRI-compatible steering wheel and pedal set. This city is composed of multiple distinct neighborhoods, traffic, and pedestrians to create a dynamic virtual environment. In the following sections, we describe each aspect of the experimental paradigm in detail.

2.2.1. CARLA

Carla is an open-source plug-in for Unreal Projects. It provides a platform for generating realistic simulated training data to train autonomous vehicle controllers. We make use of several features within Carla. These include the AI traffic system, the road generating system, the pedestrian content, and also shaders for extracting ground-truth depth and semantic segmentation information.

2.2.2. World-building

A vital component of a naturalistic navigation task is a realistic environment in which subjects can navigate. To this end, we constructed a large (2 × 3 km) virtual city (Figure 2.1). This city contains multiple neighborhoods, including urban areas, commercial zones, suburban areas, parks, and rural areas. It is highly detailed, including element such as trash bags out for collection, air conditioners in windowsills, and display shelves in stores. This detailed world is made possible by the wide availability of off-the-self 3D asset packs. We did not have to build models from scratch, a prohibitively time-consuming and difficult process that would require dedicated artists. Specifically, we made use of the PolyPixel Freebie Pack (Polypixel), Polypixel Suburbs Pack (PolyPixel), Simple City (Synty Studios), POLYGON City Pack (Synty Studios), Dynamic Street Name Signs (Nate Lauterstein), and the Shop Sign

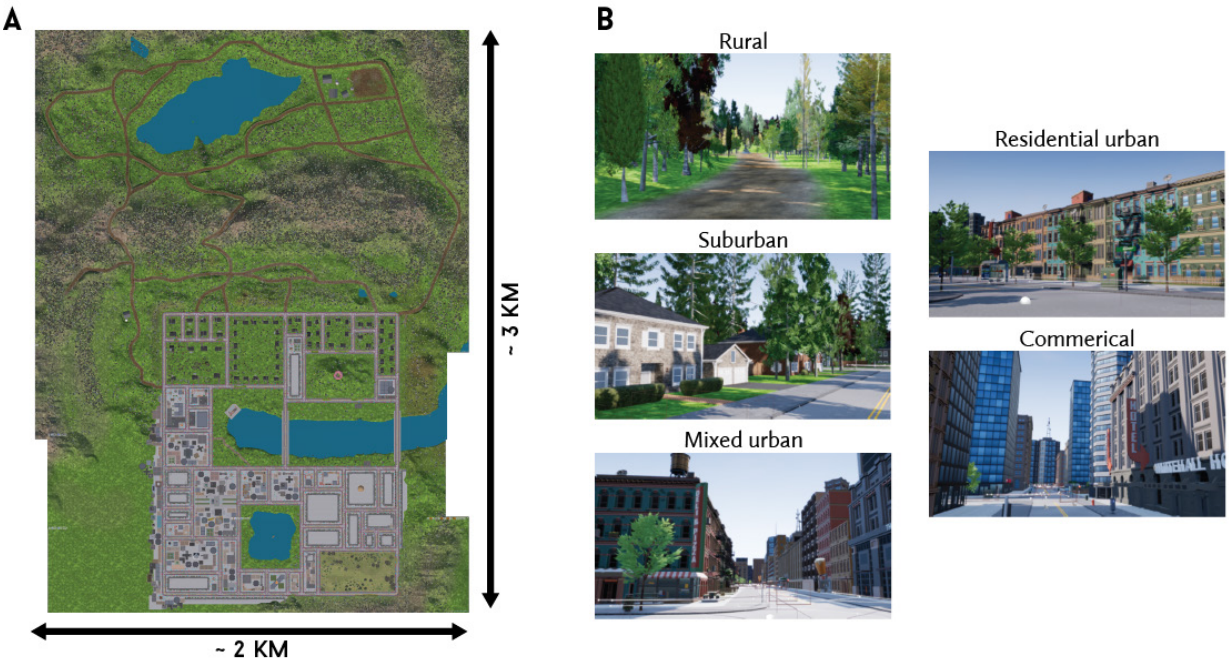


Figure 2.1 A large virtual driving environment. We used the CARLA simulator and Unreal Engine 4 to construct a large and dynamic virtual world for active navigation in the MRI. This virtual world includes weather, time-of-day lighting, and a dense AI vehicular and pedestrian traffic system. A) Top-down view of the map. The virtual world measures approximately 2 × 3 km and contains an extensive road network through which subjects drive. B) Typical views within the virtual world. The virtual world contains many different neighborhoods, including rural, suburban, mixed urban, residential urban, and commercial areas. This virtual world provides a naturalistic fMRI driving environment.

Board Pack (Raven Loh), along with many of the built-in foliage contents of UE4. These packs provide Lego-like modular components that can be combined to create virtually unlimited unique buildings.

We used the road generation system in Carla to create the road grid system for the virtual city, and augmented this road system with hand-crafted rural roads. The Carla road generation system lays out a randomly-generated road grid, and places textured meshes for the roads and sidewalks. The road generation system also creates a road map that represents the direction-of-travel at each point. Additionally, it also places scripted traffic lights at intersections. We then used built-in terrain tool in UE4 to add in hills in the rural area beyond the city.

After the creation of the road system, we manually built and placed buildings in the city using the modular content packs. The city contains over 250 uniquely identifiable locations. These include unique buildings, landmarks, named locations, and street numbers. Foliage was then added using the UE4 foliage tools. Spawn points for both vehicles and pedestrians were then placed across the map. During experiments, vehicles and pedestrians are spawned at these spawn points and create realistic traffic. These make use of the vehicle and pedestrian models provided in Carla. The vehicles are based on UE4's PhysX vehicle class, which simulates vehicle dynamics, including tire traction, suspension, transmission, and engine speed.

2.2.3. Rendering

No graphics programming was necessary, as the Unreal Engine handles all rendering. We simply specified the position of the player camera relative to the vehicle and its field of view.

2.2.4. AI

In the virtual world, AI vehicles and pedestrians move about the world to create a dynamic city. We largely used the AI provided in by Carla and UE4 with minor modifications, and provide brief overviews of their AI logic.

To create vehicles that follow traffic rules, Carla implements a custom vehicle controller class. This controller class uses the road map (see 2.2.2) to determine the direction-of-travel of the roads. On each tick, the controller checks that its vehicle is moving in the direction-of-travel, and issues corrective steering commands if needed. The controller also keeps the vehicle speed within posted speed limits. At intersections, the controller randomly picks one of the paths that exit the intersection. The controller for traffic light informs the vehicle controller of the state of the traffic light.

Pedestrians make use of the generic controller classes provided by the Unreal Engine. When a pedestrian is spawned, another random spawn point is picked as its destination. The pedestrian's controller then navigates the pedestrian to that location. To enable realistic navigation, the pedestrian controllers use a navigation mesh built for the map. In this navigation mesh, both sidewalks and roads are accessible, but roads are heavily penalized. This navigation mesh encourages the pedestrian controllers to use the sidewalks, but allows them to cross roads when necessary. Additionally, no-go zones are specified on the map. After arrival, the pedestrian is despawned, and a new pedestrian is randomly spawned. Both the vehicle and pedestrian controllers use UE4's built-in collision avoidance system, and vehicles will yield to pedestrians.

The Carla vehicle AI suffered from two shortcomings. First, because of the simplicity of the Carla vehicle controllers, AI vehicles can quickly become deadlocked in many situations, such as traffic jams in the middle of intersections or unexpected objects on the road. In such cases, vehicles are permanently stopped until the world is reset. This deadlock reduces the amount of traffic in other parts of the map. Second, because of the large size of our world, many vehicles are required to create a reasonable amount of traffic across the map. We found this to be excessively computationally intensive, as hundreds of vehicles must be simulated in real time.

To solve the first problem, we improved the vehicle controller class to despawn its vehicle if it has been immobile for more than 30 seconds. A new vehicle is then randomly spawned elsewhere to maintain a constant number of AI vehicles in the world. This improvement gradually resolves deadlocks and allows traffic to flow through the map. To solve the second problem, we improved the vehicle controller class to despawn its vehicle if it has not been rendered for more than 30 seconds. We then modified the game logic to spawn vehicles only in proximity to the subject's vehicle. Thus, AI vehicles only populate the visible environment and the immediately adjacent regions. As the subject moves across the map, AI vehicles are

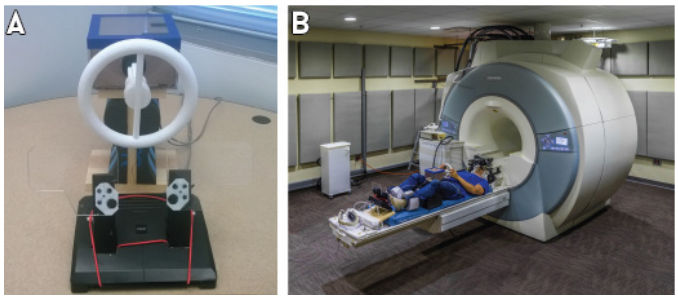


Figure 2.2 A MR-compatible steering wheel and pedal set. A) The MR-compatible steering wheel and pedal set is based on an off-the-shelf USB game controller. We removed all ferromagnetic components, including rumbler motors, springs, and sheet metal housings. These parts were replaced with custom parts 3D-printed with polylactic acid. Force feedback on the steering wheel and pedals are provided by custom mechanisms that replace torsion springs with elastic cords. Due to space constraints imposed by the scanner bore, we 3D-printed small steering wheel with a diameter of 6 inches. To improve ergonomics from a supine position, the steering wheel box is placed on a wooden platform placed between the legs. We also replaced the pedals with the larger pedals that allowed subjects to use both feet and minimize leg movement. The enclosure walls are shielded with conductive copper spray, grounded to the RF shielding of the MRI room via a shielded cable. B) A subject on the scanner bed with the steering and pedal set.

dynamically spawned around the subject to create traffic, and then despawned when not needed. This improvement allows us to simulate extremely dense traffic with fewer than 30 vehicles, significantly decreasing physics simulation time and improving framerate.

2.2.5. Experimental logic

To implement experiment logic, we subclassed the base Carla vehicle controller class and created an MRI experiment vehicle controller class. This class accepts user inputs to control the subject's vehicle, and it records TTLs from the scanner to align events with the collected brain activity. It provides a pure virtual method for handling experiment logic. We then subclass the MRI experiment vehicle controller class for each task the subject may perform. Each child class implements the experiment logic method. For example, in the taxi driver vehicle controller, this method picks destinations, generates instructions to the subject, logs progress, and checks for arrival. These controllers also provide methods for the user interface to display information. This modular inheritance structure allows new experiments and tasks to be easily implemented with minimal work.

2.2.6. User interaction

The most common method for subjects to provide feedback in the MRI are button boxes. There also exist simple two-axis joysticks and scroll wheels. They provide only binary outputs or only a few degrees of freedom in continuous outputs. These interfaces offer an extremely limited range of actions and are not suitable for a complex naturalistic task such as driving. Furthermore, they are actuated using only the fingers, which may not offer the best control precision or realism. To overcome this challenge, we modified an off-the shelf USB steering and wheel and pedal set to be MR-compatible (Figure 2.2). We largely followed a similar procedure as (Trees et al., 2014) and summarize it here.

These modifications include both MR safety, compatibility, mechanical, and also ergonomic modifications. To ensure MR safety, we removed all ferromagnetic components. These components included the rumbler motors, screws, springs, and the housing of the steering wheel. The original screws were replaced with brass screws, and the housings were replaced with custom 3D-printed ABS parts and laser-cut PMMA panels that were joined with epoxy. We did not replace the rumbler motors, as force feedback was not important to our experiment.

To ensure MR-compatibility, we added electromagnetic shielding to the housings to protect the controller boards from the RF from the MR gradients. We used Super Shield silver coater conductive paint (MG Chemicals) on the inside of the housings. The standard USB cables were replaced with Belden 8723 shielded cables (Belden Inc). These shielded cabled provided an extra grounding wire that we connected to the EM shielding layer on the inside of the housing. The cable passed through the shielding of the MR room via a d-sub connector, and then converted back to a USB interface outside the scanner room. The ground wire was connected to the housing of the d-sub connector, which in turn is electrically connected to the shielding of the scanner room. Thus, the interior of the steering wheel could be electrically considered a protrusion of the unshielded space into the closed scanner room.

Mechanically, we fabricated replacements for the springs that were removed. Three torsion springs provided mechanical resistance in the steering wheel and pedals, and also served to return the wheel and pedals to their neutral positions. These springs were replaced with tensioned bungee cords.

We added anchor points to the arms of the pedals, and attached the bungee cords to these anchors. The other ends of the bungee cords were anchored to points in front of the pivots for the pedals. These tensioned bungee cords provide resistance to pedal depression. To replace the spring that provided torsional resistance to the steering wheel, we designed a ring with extended arms that slipped over the steering wheel shaft. The ring locked on to the ridges in the steering wheel shaft, and two arms extend symmetrically outward from the ring. Two bungee cords were then attached to the arms. The other ends of the cords were then wrapped across the steering wheel shaft and anchored into the housing. One cord was wrapped clockwise and the other was wrapped counterclockwise, and both were equally tensioned. Thus, the cords were equally tensioned at the neutral position, and any rotation of the wheel decreased tension in one arm and increased tension in the other, providing resistance.

Finally, we ergonomically modified the steering and pedals to be better suited to a supine position in the scanner rather than a sitting position. We built an I-shaped mount for the steering wheel housing, so that it can be placed between the subject's legs. We then 3D printed a shaft extension, so that the steering wheel is reachable when the housing is placed between the legs. We also 3D printed a smaller diameter steering wheel that better fit into the scanner bore. The pedal housing was also raised on an angled mount, such that in their neutral position, the pedals were vertical with respect to the bed surface. The pedals were also replaced with wider pedals cut from PMMA sheets. Thus, the pedals could be actuated with one foot each to minimize motion.

2.2.7. MR-compatibility validation

We validated that this steering wheel and pedal set did not adversely affect image quality. First, we used a phantom to ensure that the controls did not introduce artifacts in the image when placed at the normal usage position. Second, and more importantly, we examined the effects on functional SNR in a realistic fMRI paradigm.

2.2.7.1. AVSNR task

A subject watched 8 repeats of a 2-minute audio-visual stimulus at fixation while BOLD activity was recorded. This video was designed to measure the audio-visual signal-to-noise-ratio (AVSNR) in the brain. In 4 runs, the subject watched the video without the steering wheel and pedal set in the scanner. In the other 4 runs, the subjects watched the video with the steering wheel and pedals, and also actuated the steering wheel and pedals to simulate driving.

2.2.7.2. Scanning procedure

MRI data were acquired on a 3T Siemens Trio at the University of California, Berkeley with a 32-channel head coil. BOLD data were acquired using a T2*-weighted gradient-echo EPI sequence customized with a water-excitation radiofrequency pulse to prevent contamination from fat signal. We used the following parameters: TR = 2 s, echo time = 34 ms, flip angle = 74°, voxel size = 2.24 × 2.24 × 3.5 mm^3, field of view = 224 × 224 mm^2, matrix size = 100 × 100, and 32 axial slices to cover the entire cortex. A custom personalized headcase (Power et al., 2019) was used to stabilize the head and reduce motion artifacts. To reconstruct cortical surfaces, we collected anatomical data with 1 × 1 × 1 mm^3 voxel size and 256 × 212 × 256 mm^3 field of view using a three-dimensional T1-weighted MP-RAGE sequence.

2.2.7.3. Explainable variance calculation

In each condition, we computed the explainable variance (EV) across the 4 repeats as a measure of functional SNR. EV is defined as the fraction of variance that the mean timecourse across repeats is able to explain in the data. EV assumes that only repeatable signal is true functional signal, and all other variance across runs is noise. In other words, EV assumes that the functional signal is purely stimulus-driven. For each voxel, we then compared the EV between the two conditions.

2.2.8. Feature extraction

Ground truth about the stimulus, subject behavior, and events in the virtual world are necessary for modeling brain activity. Such information includes the visual scene as seen by the subject, semantic segmentation of that scene, the control inputs, the state of the traffic lights, the position of other vehicles and pedestrians, among others. Currently, we extract 33 different types of features, ranging from low-level visual information to abstract navigational representations. A full list can be found in section 3.2.9 Features.

In non-interactive experiments, feature timecourses are determined beforehand during stimulus design or at the start of trials. In interactive experiment such as this one, this information cannot be predetermined. Actions taken by the subject directly affect the stimulus shown. No two subjects take the exact same actions at the same time, and a subject cannot perfectly replicate previous actions. Thus, the exact stimulus features are unique to every run in every subject. Features must therefore be extracted from every run.

We cannot to extract all such features on-line during the experiment due to their computational demands. Here, we define game time as time elapsed in the virtual world. During the experiment, game time must progress at the same rate as real time to ensure that the subject can interact seamlessly with the virtual world. Feature extraction is computationally intensive, and causes game time to flow slower than real time. We instead perform feature extraction offline after the experiment, which requires producing an exact record of the experiment. We do so with OBS Studio (OBS Project), an eyetracking camera (Avotec), and UE4's built-in game recording system.

During the experiment, we use OBS studio to capture the video and audio output at 30 fps. The recordings produced by OBS studio accurately reflects what the subject sees and hears during the experiment. Simultaneously, we record gaze location at 60 fps with the eyetracker. Finally, we use UE4's demo recording system efficiently and accurately record all events in the world. These events include experimental events, such as cueing of a new destination, the controls inputs from the subject and the current gear selected by the subject, and also all world events, such as traffic light states, positions of other vehicles, and collisions between vehicles. The recording system produces a demo file, which use Unreal Engine's multiplayer networking system to efficiently compresses world events. These records are accurate, as multiplayer servers must produce an absolute veridical account of events to resolve conflicts due to network latency.

After the experiment, we then play back the demo files to completely reconstruct the events from the experiment and extract features. Because offline feature extraction does not need to be interactive, we decouple game time from real time and step through the recording at a constant game time step of 15 fps. On each frame, we extract all features before moving on to the next time step. This feature extraction process can be re-run if needed to extract further features, or at different framerates, making this system extremely flexible. This system even allows retroactive extraction of features that had not been designed or conceived at the time of the experiment.

2.2.9. Deployment

The driving simulator is designed to run on multiple client machines in varied settings. The three key use cases are presenting stimuli during the experiment, training subjects outside the scanner, and extracting features after an experiment. In the first and third case, we directly manage the computer running the simulator. However, in the second case, subjects typically practice at home on their own machines, which we cannot manage. Furthermore, the development environment totals over 60 GB. This environment includes the source code,

unoptimized 3D assets, the game engine, and game editor. Both unmanaged machines and the size of the codebase present challenges to deploying and updating the driving simulator on client machines.

To facilitate deployment, we developed a cloud-backed solution. First, the driving simulator is compiled to a self-contained executable. This is in contrast with most experimental stimulus presentation frameworks, such as those based in python or Matlab, that are dependent on the existence of a local interpreter and libraries. The driving simulator is then uploaded to a cloud-based storage service, along with versioning information. In our case, we use Wasabi, a s3-compatible cloud-based object storage provider. Second, we built a lightweight launcher program (~100 mb) that we manually distribute to managed client computers and to subjects. We also upload the launcher, with its own versioning information, to the cloud-based object store. On the client machine, the launcher polls cloud storage for the latest version of the simulator, and automatically downloads and updates the local copy. The launcher, via an included updater, also updates itself to the latest version.

On each build of the driving simulator, we simply upload the new version to the object store. The launcher on all client machines will then automatically update the deployed version. This system greatly reduces the management and deployment workload, while simultaneously ensuring that deployed instances of the simulator are up-to-date with no active management.

2.3. Results

As part of the realistic driving simulator, we constructed a set of MR-compatible steering wheel and pedals set. Because this controller set includes electronics in the MRI room, we examined its effects on functional SNR. A subject watched multiple repeats of a video both with and without the steering wheel and pedals set. We then computed EV across the brain in both conditions.

We find that the distribution of EV across the brain is largely comparable across the two conditions (Figure 2.3A). We then examined the distribution of well-predicted voxels. We picked an arbitrary threshold of 0.4 EV in at least one of the conditions to define as well-predicted. In this set of voxels, 3085 had higher EV in the no-steering wheel condition, and 2494 had higher EV in the condition with steering wheels.

To better understand this discrepancy, we plotted the EV in both conditions on the cortical surface (Figure 2.3B). We find that voxels with higher EV in the no-steering wheel condition are largely concentrated in the anterior IPS and the somatosensory and motor cortices. On the other hand, voxels in visual and auditory cortex generally displayed similar EV across conditions.

We explain this difference by noting that EV assumes that only replicable signal across repeats is true signal, and any variance is noise. In the condition without the steering wheel and pedals, the subject simply watched the video at fixation. Thus, brain activity in this condition should predominantly be bottom-up stimulus-driven. In the condition with steering wheel and pedals, the subject also actuated the controls to better mimic a

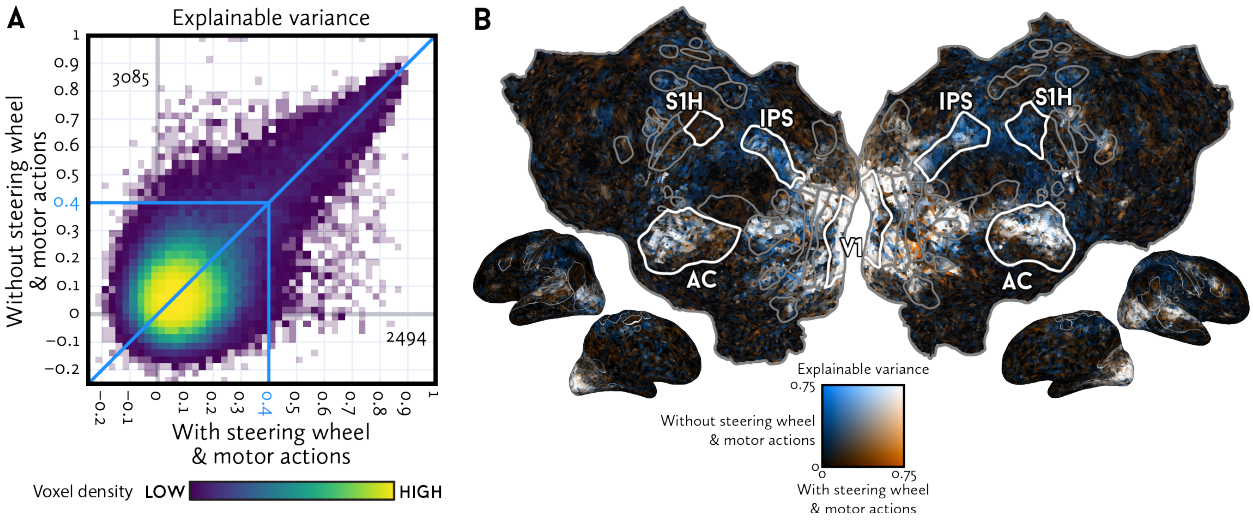


Figure 2.3 We examined effects from the steering wheel and pedal set on functional SNR in a realistic fMRI paradigm. A subject watched 8 repeats of a 2-minute video with audio while BOLD activity was recorded. In 4 runs, the subject watched the video without the steering wheel and pedal set in the scanner (without-steering wheel). In the other 4 runs, the subject watched the video with the steering wheel and pedals, and also actuated the steering wheel and pedals to simulate driving (with-steering wheel). In each condition, we computed the explainable variance (EV) across the 4 repeats as a measure of functional SNR. A) 2D histogram of voxel EV across the conditions. Brighter colors correspond to more voxels in each histogram bin. Voxels along the diagonal display comparable EV across conditions, while voxels above the diagonal show higher EV in the without-steering wheel condition, and vice-versa for voxels below the diagonal. In the set of voxels with a minimum EV of 0.4 in at least one of the conditions, 3085 voxels had higher EV in the without-steering wheel condition, and 2494 had higher in the without-steering wheel condition. B) We plot EV in the two conditions on the cortical surface. Blue voxels show higher EV in the without-steering wheel condition, orange voxels show higher EV in with-steering wheel condition, and white voxel show comparable EV in both conditions. We note that EV assumes that only replicable signal is true signal. In the with-steering wheel condition, in addition to the video, the subject must also attend to and actuate the steering wheel and pedals. The subject cannot perfectly replicate these internally-driven attention and motor signals across repeats. EV cannot capture this non-replicable signal and considers it noise. The reduction in EV in the with-steering wheel condition is largely confined to visual attention and somatomotor areas. These regions are sensitive to top-down signals, and the loss of EV reflects its inability to capture non-replicable signal, rather than a loss of SNR. On the other hand, EV is comparable across conditions in more stimulus-driven areas such as the primary visual and auditory cortices. Because EV is consistent across conditions in stimulus-driven regions, we conclude that the steering wheel and pedals set do not significantly affect functional SNR.

navigation experiment. Because the subject must think about and actuate the controls while also watching the stimulus, brain activity in this condition also contains a significant top-down internally-driven component. However, the subject cannot perfectly replicate these internally-driven signals between repeats of the same stimulus. Thus, some true signal varies across repeats and is treated by the EV metric as noise. This leads to an apparent drop in SNR as measured by this metric.

The loss in EV in the steering wheel condition is largely confined to visual attention and motor areas. These regions are more sensitive to top-down signals, and their activity are thus more likely to vary across repeats in the steering wheel condition. The EV metric is thus ill-suited for capturing functional SNR in these regions. On the other hand, perceptual

areas such visual and auditory cortex show similar EV in the two conditions. The activity in these regions is more stimulus-driven and are thus more likely to be consistent across repeats regardless of top-down signals. The EV metric thus more accurately captures the functional SNR in these regions. Indeed, if the steering wheel and pedals do indeed affect true functional SNR, its effects should be global or be delineated by spatial volumes or slices, rather than be delineated by functional brain regions. Thus, because EV is consistent across conditions in more perceptually-driven regions, we conclude that the steering wheel and pedals set do not significantly affect functional SNR.

2.4. Discussion

A significant challenge to fMRI studies of human navigation is creating an ecologically valid, active navigation task that emulated real-world navigation. Previous studies have mainly used tightly controlled stimuli and passive tasks that are not representative of the demands of naturalistic navigation. Here, we have overcome this issue with a new, active navigation paradigm. This paradigm leverages the capabilities of Unreal Engine 4, a state-of-the-art game engine, to create a large, dynamic, and naturalistic virtual world in which subjects can actively navigate by driving. We designed the software architecture of the virtual world to be modular and flexible, so that new experiments can be easily added.

We also built a MR-compatible steering wheel and pedal that offers a realistic driving experience, as opposed to simple button box or joystick controls commonly used in fMRI experiments. The electronics in the steering and pedal set do not generate RF interference or affect data quality. Subjects find this world to be realistic and immersive, and the controls to be easy to use. This framework offers an unparalleled environment in which to study active naturalistic navigation with fMRI.

Currently, this framework is focused on the study of navigation. However, this paradigm could be extended to the study of many other cognitive tasks. Like navigation, many tasks are inherently interactive. For example, social tasks necessarily require interacting with other people, which is difficult to emulate in traditional experimental designs. The Unreal Engine provides a general programming framework for building many different interactive tasks. Indeed, the Unreal Engine is used for many types of games and applications, including shooters, platformers, open-world games, dynamic simulations, and computer-generated imagery for movies and film. We could easily implement non-navigational tasks in our current world, and new interactive experiments are straightforward to implement in Unreal Engine.

The current control interface, i.e., the steering wheel and pedal set, is designed specifically for navigation via driving. To provide a more general interface, we have also constructed a MR-compatible video game controller by applying the same modification process to an off-the-shelf Xbox controller. This controller will allow subjects perform many more tasks. This framework will enable the study of many more complex and naturalistic tasks with fMRI.

3 Mapping navigational representations across the brain during naturalistic driving

3.1. Introduction

Previous experiments have revealed a complex network of brain regions that are engaged during navigation. This network includes the parahippocampal place area (PPA) (Epstein et al., 2003; Epstein and Kanwisher, 1998), occipital place area (OPA) (Dilks et al., 2013; Grill-Spector, 2003), retrosplenial cortex (RSC) (Maguire, 2001; Vass and Epstein, 2013), parts of the prefrontal cortex (PFC) (Brown et al., 2016; Granon and Poucet, 1995; Hok et al., 2005; Rossi et al., 2012; Yoon et al., 2008), the hippocampus (Burgess et al., 2002; Maguire et al., 2006; O'Keefe and Dostrovsky, 1971), and entorhinal cortex (EC) (Doeller et al., 2010; Fyhn et al., 2004; Sargolini et al., 2006). This network perceives the world around the subject, maps this world on to an abstract allocentric map, and uses these representations to plan and navigate through the world. Each of the regions in the navigational network may represent multiple types of related navigational features. However, the vast majority of human neuroimaging studies isolated individual regions and representations from rest of the navigational network. The tasks used in most studies also often do not reflect the full demands of active navigation. The result is many piecemeal insights into how the human brain supports navigation. An integrated description of the representation of navigational information, particularly during active navigation, remain elusive.

Previous studies have been limited by two critical barriers: tightly controlled stimuli and passive tasks that are not ecologically valid for studying navigation, and statistical methods that cannot analyze high dimensional data from complex tasks. Most experiments use

tightly controlled stimuli to isolate individual hypotheses. However, because the brain is a nonlinear dynamical system, it behaves differently under simplified conditions (Cremers et al., 2017; Desmond and Glover, 2002; Hasson et al., 2010; Matusz et al., 2019; Mumford and Nichols, 2008; MC-K Wu et al., 2006). Thus, the results from controlled experiments may not generalize to navigation in the real world. The statistical methods used in most fMRI studies, such as statistical parametric mapping (Friston, 1994b), representational similarity analysis (Kriegeskorte et al., 2008), or multivariate pattern analysis (Norman et al., 2006), are designed to test individual hypotheses and cannot examine many possible representations simultaneously. These methods cannot capture brain representations in a complex, dynamic, and open-ended task such as active navigation. Thus, these previous experiments cannot fully capture and map multiple navigational representations across the brain.

In this experiment, we overcome these limitations with a naturalistic active navigation experiment. We developed an experimental framework using Unreal Engine 4 and MR-compatible controls (see section 2.2). Subjects actively navigate in a large 2×3 km virtual world while brain activity is recorded with fMRI.

We then apply voxelwise modeling (Çukur et al., 2013; Deniz et al., 2019; Huth et al., 2012, 2016; Lescroart and Gallant, 2019; Naselaris et al., 2011; Nishimoto et al., 2011) with banded ridge regression (Nunez-Elizalde et al., 2019) to simultaneously map representations of 33 different navigational feature spaces across the cortex. As opposed to previous experiments that focused on specific hypotheses about brain regions, this data-driven experiment allows us to discover novel brain regions and networks related to navigational representations.

3.2. Methods

3.2.1. Subjects

Three healthy adult volunteers (1 female) with normal or corrected to normal vision participated in this study: S1 (age 26), S2 (age 27), and S3 (age 32). The experimental procedures were approved by the Institutional Review Board at the University of California, Berkeley, and written informed consent was obtained from all subjects.

3.2.2. Virtual environment and driving simulator

Using Unreal Engine 4 and the Carla plugin (Dosovitskiy et al., 2017), we built a driving simulator in 2×3 km virtual city complete with dynamic vehicular traffic and pedestrians (see 2.2). This city contains multiple neighborhoods, including urban, commercial, suburban, and rural areas, each with a distinct visual appearance. The map contains over 200 uniquely identifiable locations. AI vehicles obey traffic lights and posted speed limits, and respect pedestrian right-of-way. This provides an ecologically valid, dynamic world in which subjects can actively navigate by driving a virtual car. Video is projected into a screen in the bore. Subjects are shown a first-person view from the virtual car, and drive using a set of MR-compatible steering wheel and pedals. Subjects can also use a button to toggle between forward and reverse gears.

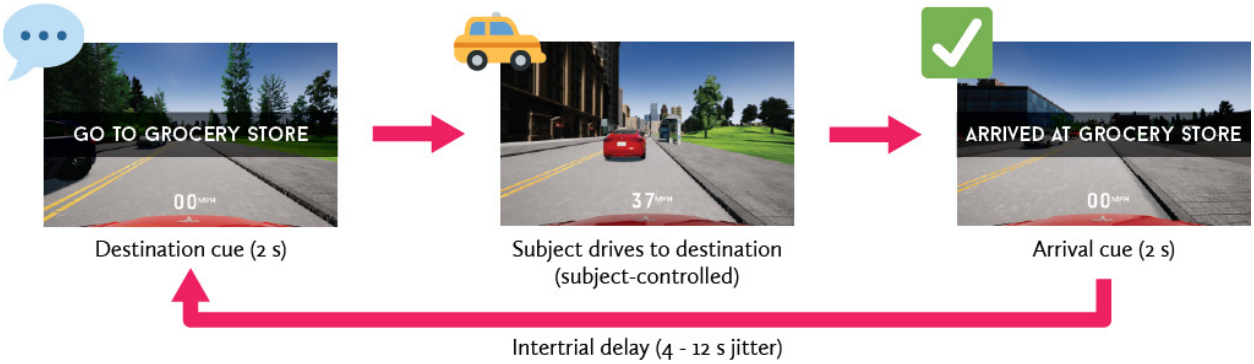


Figure 3.1 Subjects performed an active driving task in the MRI. Prior to being scanned, subjects learned a set of 77 possible locations in the map. In the scanner, the subject is successively cued to drive to randomly selected locations. On each trial, a random destination is selected, and the destination is shown on the screen for 2 seconds. Subjects the drive to the destination via the fastest path, while navigating through traffic and obeying all traffic rules. At the destinations, subjects must come to a complete stop to indicate that they have arrived at the destination. An arrival cue is then shown to acknowledge the end of the trial. After a jittered delay of between 4 to 12 seconds, a new trial begins with another randomly selected destination.

Using the built-in Unreal Engine recording system and OBS studio, the driving simulator produces a complete record of the experiments. This record includes ground truth about the local environment from the subject's point of view, the control inputs and behavior of the subjects, the navigational goal of the subject, and the behavior of all other vehicles and pedestrians. Additionally, eye motion is recorded continuously at 60 Hz to facilitate data analysis and modeling. These records are then used to extract features from the experiments offline.

3.2.3. Task

In this world, subjects perform a taxi-driver task (Figure 3.1). We selected 77 locations spaced across the virtual world, and subjects are directed to drive to these destinations. These locations included landmarks (e.g., a statue of a fire hydrant), special named locations (e.g., the Midtown Bank), house numbers (e.g., 1 Broadway). Prior to the experiment, subjects learnt the world outside the scanner.

The experiment consists of many navigational trials in which the subjects navigate to a destination. Each trial is unique, as the route and duration are decided on-the-fly by the subject. At the beginning of each trial, a new destination is randomly selected. A text cue for the destination is displayed at the center of the screen for 2 seconds at the center of the screen. Subjects then drive to the destination via the quickest path. When subjects arrive at the destination, they must stop completely to indicate that they have arrived. Merely passing by the destination does not constitute arrival. This ensures that the subjects are deliberately navigating to the goal, and precludes them from blindly searching. A text prompt is then displayed for 2 seconds to acknowledge their arrival. After a jitter of 4-12 seconds, a new trial begins. Subjects navigate at their own pace and must respond to the dynamics of the world. Subjects are allowed to freely view the world. (In pilot experiments, we found that it was difficult to drive while fixating.)

Note that due to its open-ended nature, the task could continue indefinitely. In pilot experiments, we found that approximately two hours of data were necessary to fit models. Thus, subjects performed this task for a minimum of two hours.

3.2.4. Scanning procedure

MRI data were acquired on a 3T Siemens Trio at the University of California, Berkeley with a 32-channel head coil. BOLD data were acquired using a T2*-weighted gradient-echo EPI sequence customized with a water-excitation radiofrequency pulse to prevent contamination from fat signal. We used the following parameters: TR = 2 s, echo time = 34 ms, flip angle = 74°, voxel size = 2.24 × 2.24 × 3.5 mm^3, field of view = 224 × 224 mm^2, matrix size = 100 × 100, and 32 axial slices to cover the entire cortex. Custom personalized headcases (Power et al., 2019) were used to stabilize the head and reduce motion artifacts. To reconstruct cortical surfaces, we collected anatomical data with 1 × 1 × 1 mm^3 voxel size and 256 × 212 × 256 mm^3 field of view using a three-dimensional T1-weighted MP-RAGE sequence.

Data was collected in 11-minute runs. We collected 11 runs of data from subject 1, and 18 runs of data each from subjects 2 and 3.

Eyetracking data was collected using an Avotec dark-pupil IR eyetracker at 60 Hz, and processed with custom software to extract gaze locations. At the beginning of every functional run, 35 calibration points were presented for 2 seconds each to subjects to ensure accurate eyetracking calibration. The taxi driver task began after the calibration sequence.

3.2.5. fMRI data preprocessing

Each functional run was first motion-corrected using the FMRIB Linear Image Registration Tool (FLIRT) from FSL 5.0 (Jenkinson et al., 2002; Jenkinson and Smith, 2001). Then, FUGUE from FSL was used to unwarp the functional images using fieldmaps collected between functional runs. All volumes in the run were then averaged across time to obtain a high-quality template volume. The template from the first functional run was chosen as the overall template to which all functional runs were aligned. The functional data were then aligned to the anatomical surface with pycortex (Gao et al., 2015). These automatic alignments were manually checked and adjusted as necessary to improve accuracy. Low-frequency voxel response drift was identified using COMPCOR (Behzadi et al., 2007) and removed from the signal. Physiological signals from breathing and heartbeats were also regressed out. Voxel activity in each run were z-scored separately; that is, the mean response for each voxel was subtracted and the remaining response was scaled to have unit variance. Then, the first 35 and last 5 TRs were discarded from each run to remove scanner and detrending artifacts. We discard a larger number of TRs at the beginning to remove data from the eyetracking calibration sequence, which is irrelevant for navigation.

3.2.6. Cortical surface reconstruction

Freesurfer (Dale et al., 1999) was used to reconstruct cortical surface meshes from the T1-weighted anatomical volumes. The anatomic segmentation was checked by hand, and manually corrected where necessary with Blender (Blender Foundation) and pycortex.

Blender and pycortex were then used to remove the medial wall, and relaxation cuts were then made into each surface. The cut at the calcarine sulcus was made using retinotopic localizers as a guide to bisect V1 along the horizontal meridian.

3.2.7. Localizers for known ROIs

Known ROIs were localized in each subject in separate experiments. ROIs were determined using five sets of localizers: a retinotopic localizer, a MT localizer, a visual category localizer, a motor localizer, and a theory-of-mind localizer.

3.2.7.1. Retinotopic Localizer

Retinotopic localizers were collected in four 9-minute runs. In two runs, rotating wedges were used to determine visual angles. In the other two runs, expanding rings were used to determine eccentricity. Visual angle and eccentricity maps were then used to delineate V1, V2, V3, V4, V3A, V3B, and V7 (Hansen et al., 2007).

3.2.7.2. MT Localizer

MT localizer data were collected in four 90-second runs. These runs consisted of alternating 16-second blocks of motion coherent and incoherent random dot fields. The contrast between coherent and incoherent blocks was used to delineate MT+.

3.2.7.3. Visual category localizer

Visual category localizer data were collected in six 4.5-minute runs. These runs consisted of 16 16-second blocks. In each block, 20 images of either places, faces, body parts, non-human animals, household objects, or phase-scrambled objects were displayed. Each image was displayed for 300 ms followed by a 500 ms blank. The contrast between faces and objects was used to define the fusiform face area (FFA) (Kanwisher et al., 1997). The contrast between body parts and objects was used to define the extrastriate body area (EBA) (Downing et al., 2001). The contrast between places and objects was used to define the parahippocampal place area (PPA) (Epstein and Kanwisher, 1998), occipital place area (OPA) (Dilks et al., 2013), and retrosplenial cortex (RSC) (Aguirre et al., 1998).

3.2.7.4. Motor localizer

Motor localizer data were collected in one 10-minute run. Subjects were cued to randomly alternate between six motor tasks in 20-second blocks. For “hand,” “foot,” “mouth,” “speech,” and “rest,” the cue was simply that word presented at the center of the screen. For the saccade condition, subjects were shown a screen filled with randomly distributed saccade targets.

In the “hand” blocks, subjects were instructed to make small finger-drumming movements. In the “foot” blocks, subjects were instructed to make small toe movements. In the “mouth” blocks, subjects were instructed to make small lip movements without moving the tongue or jaw. In the “speak” blocks, subjects were instructed to continuously subvocalize self-generated sentences without physically speaking. In the “saccade” blocks, subjects were

instructed to continuously saccade between targets presented on the screen. A linear model was used to find the change in BOLD response in each voxel in each condition relative to the mean BOLD response.

Weights for the hand, foot, and mouth responses were used to delineate the primary motor and somatosensory areas for hands (M1H, S1H), feet (M1F, S1F), and mouth (M1M, S1M). Weights for the saccade responses were used to delineate the intraparietal sulcus (IPS), frontal eye fields (FEF) (Paus, 1996), and frontal operculum (FO) (Corbetta et al., 1998). Weights for the speak responses were used to delineate Broca's area and the superior ventral premotor speech area (sPMv).

3.2.7.5. Theory-of-mind localizer

Theory of mind localizer data were collected in two 5-minute runs. Subjects were shown a series of short stories consisting of false belief, false photograph, desire, physical description, and nonhuman description stories (Saxe and Kanwisher, 2003). Each story was shown for 10 seconds. After each story, subjects were presented with a fill-in-the-blank two-alternative forced choice question for 4 seconds and responded with a button box. Half of the questions concerned the belief of characters in the stories, while the other half were concerned factual information. The contrast between belief and factual questions was used to delineate the temporoparietal junction (TPJ).

3.2.8. Behavioral measure for navigation

Subjects were asked to self-report that they were familiar with the environment prior to scanning. However, navigational ability varies significantly between individuals. Training subjects to improve navigational skills is beyond the scope of this study. Instead, we quantify the ability of subjects to navigate with a *route optimality* metric, and establish a cutoff criterion to determine when subjects have learned the map.

The *route optimality* metric is defined as follows. For each trial, we first calculate the total distance traversed by the subject to reach the destination. Next, we use A* to search through a navigation mesh for an optimal path between the starting location and the destination. This navigation mesh is the inverse of the navigation mesh specified in 2.2.4 AI for pedestrian AI navigation. Non-road areas are heavily penalized while roads are lightly penalized to encourage A* to search for routes that traverse only roads. Because A* is a heuristic search and not guaranteed to find a global optimal path, we perform the search a second time, starting from the destination and ending at the starting point of each trial. Note that this is not the same as bidirectional A*, in which the search starts simultaneously from the start and end, and meet in the middle. We then take the shorter of the two as the optimal path. Route optimality is then defined as the ratio between the distance traversed by the subject and the length of this optimal path.

A route optimality of 1 indicates that the subject is a perfect navigator, and values above 1 indicate suboptimal route choices. Because the navigation mesh is a discretization of the map, there are edge cases in which the optimal path may be slightly longer than the actual path taken by the subject. We therefore apply a floor of 1 to route optimality. Note in the

Table 3.1 List of feature spaces developed for voxelwise modeling in the navigation experiment, and a brief description of teach feature space.

Feature Space	**Description**
Motion-energy (frame-wise)	Spatialtemporal gabor energy in frame coordinates
Motion-energy (retinotopic)	Spatialtemporal gabor energy in retinotopic coordinates
Eyetracking	Gaze location in screen space using Cartesian coordinates
Depth	Distance to objects in the virtual world
Frame Semantics	Semantic content of the screen
Gaze Semantics	Semantic content of a 5° circle around the gaze location
Spatial semantics	Semantic content of virtual world around the subject
Frame Scene structure	Spatial distribution of surfaces and their orientations around the subject in frame coordinates
Controls	Steering, gas, brake, and gear selections from the subject
Affordances	Where in the scene that the subject can legally drive
Route Progression Phase (Spatial)	Fraction of total distance traveled to the destination at any point in time, encoded by a Fourier basis
Route Progression Phase (temporal)	Fraction of total time to the destination at any point in time, encoded by a Fourier basis
Binned route progression (space)	Fraction of total distance traveled to the destination at any point in time, encoded by a set of indicator features
Binned time progression (temporal)	Fraction of total time to the destination at any point in time, encoded by a set of indicator features
Turns (spatial)	Fraction of total distance traveled until the next turn
Turns (temporal)	Fraction of total time traveled until the next turn
Destination Vector	An egocentric vector to the current destination
Destination-anchored vector	An allocentric vector from the destination to the current location of the subject
Grid cells	Position of the subject in the world expressed using a hexagonal lattice basis
Beeline distance elapsed	Beeline distance from the beginning of the current route
Beeline distance remaining	Beeline distance to the destination
Path distance elapsed	Actual distance traveled from the beginning of the route
Path distance remaining	Distance remaining along the paths to the destination
Egocentric path integration	Egocentric vector to the beginning of the current route
Allocentric path integration	Allocentric vector to the beginning of the current route
Vehicles	Positions of other cars in egocentric coordinates
Pedestrians	Positions of pedestrians in egocentric coordinates
Future path	Future path that the subject will take in egocentric coordinates
Road graph	Which edges on a road graph form the planned route
Head direction	Current allocentric direction that the subject is facing
Gaze direction	Current allocentric direction that the subject is looking (note this can differ significantly from head direction)
Gaze grid	Gaze position using a hexagonal lattice basis
Destination grid representation	Position of the destination expressed using a hexagonal lattice basis

exception of these edge cases, all trials will have a route optimality above 1 because actual paths may be noisy, or subjects may have used other heuristics to navigate. Additionally, the A* search does not respect the direction-of-travel on roads. Thus, there can be cases in which the "optimal" path is significantly shorter than the path that a driver is able to take. We required subjects to reach a median route optimality of 1.25 or lower before data collection.

3.2.9. Features

After scanning, we extracted features offline to use with voxelwise modeling. We used both the animal and human navigation literature to define navigational feature spaces. In addition to literature-derived feature spaces, we also included other feature spaces that we hypothesized may be navigationally informative but has not been investigated in the literature. We extracted 33 features spaces that encompass many aspects of navigation, from low-level visual features to high level abstract navigational representations. The list of features and a brief description of each feature space can be found in Table 3.1. In the following sections, we provide more in-depth descriptions of how each feature space is constructed. Features were computed at 15 fps unless otherwise specified.

3.2.9.1. Motion-energy (frame-wise)

The motion-energy feature space consists of a pyramid of spatiotemporal Gabor filters (Nishimoto and Gallant, 2011). These filters extract low-level visual features from screen during driving. First, screen recording videos were downsampled from 1024 × 768 to 128 × 96 pixels. Next, the color frames were converted to grayscale. The spatiotemporal Gabor pyramid were created using 7 spatial frequencies (0, 1, 2, 4, 8, 16, and 32 cycles across the image), 3 temporal frequencies (0, 2, and 4 Hz), and 8 orientations (0, 45, 90, 135, 180, 225, 270, and 315 degrees). These filters tiled the screen, and the video frames were convolved with these filters. Finally, filter responses were squared and summed for each quadrature pair for a total of 8418 features. These features were computed at 30 fps.

3.2.9.2. Motion-energy (retinotopic)

Because subjects were not fixating, we also computed motion-energy filter responses in retinotopic coordinates. First, for each frame in the screen recording videos, we found the corresponding gaze position using the eyetracking data. Then, each frame was placed into a 256 × 192 uniform gray (12.5% brightness) background such that the gaze position is centered in this larger frame. Finally, motion-energy features were computed using these frames with the same methods for computing frame-wise motion-energy features for a total of 8418 features. These features were computed at 30 fps.

3.2.9.3. Eyetracking

From the eyetracking videos, we determine the gaze position in screen coordinates. We set the origin of these coordinates to (512, 384), i.e., the center of the screen. For each dimension (x and y), we also compute the second and third powers and the first, second, and third

derivatives of the position values. Additionally, we also include the conversion of these cartesian coordinates into polar coordinates. We then rectify each channel for a total of 28 features. These features were computed at 60 fps.

3.2.9.4. Depth

The depth feature space captures the distribution of distances to objects in the virtual world. To compute the depth features, we first use the demo files to reconstruct the experiment. Next, in this reconstruction, we take the player camera, and apply a depth postprocessing effect that determines the distance to the surface rendered at each pixel. Distances are encoded with a 24-bit floating point representation up to 1000 m. This encoding provides a precision of 0.06 mm. Then, the frame is divided up into 5 horizontal bins that each span 22 degrees in the virtual environment. Six distance bins at 0-5 m, 5-10 m, 10-25 m, 25-50 m, 50-100 m, and 100+ m were created. In each horizontal bin, we then counted the number of pixels that fell within each distance bin for a total of 30 features.

3.2.9.5. Frame semantics

The frame semantics feature space captures the visual semantic content of the entire screen. To compute the frame semantic features, we again used the reconstruction of the experiment. We apply a semantic segmentation postprocessing effect to the subject's view to determine the category of the object rendered at each pixel. We used 16 categories that were relevant to driving: buildings, pedestrians, vehicles, roads, road lines, sidewalks, traffic signs, poles, walls, fences, fields, ground, foliage, self, miscellaneous objects, and sky. For each frame, we determined the fraction of the frame occupied by each category for a total of 16 features.

3.2.9.6. Gaze semantics

Subjects are unlikely to allocate equal attention to all areas of the screen. Thus, the frame semantics may not accurately reflect representations of visual semantics with the influence of top-down attention. We therefore also created a set of visual semantic features for the part of the screen at which the subject is fixating. We first took the same semantic segmentation frames used to create the frame semantic features (see 3.2.9.5). Then, for each frame, we selected a circle of 5° diameter around the fixation location. We then computed the fraction of the circle occupied by each category for a total of 16 features. These features were computed at 60 fps.

3.2.9.7. Spatial semantics

The three previous visual semantic feature spaces are all computed in the 2-dimensional image space of the eye or screen. To be navigationally useful, these semantics must be converted into the 3-dimensional physical space of the virtual world. We built the spatial semantic feature spaces to capture the 3D spatial distribution of visual semantics around the subject in the virtual world. First, we combine the depth and semantic segmentation information to create a combined distant-and-object-type label for each pixel. Then

we divide the frame up into 7 horizontal bins, each spanning 15.7 degrees in the virtual environment, and 6 distance bins at 0-5 m, 5-10 m, 10-25 m, 25-50 m, 50-100 m, and 100+ m. Finally, for each combination of horizontal and distance bins, we count the number of pixels in each semantic category in that bin for a total of 672 features.

3.2.9.8. Frame scene structure

To capture the geometry of the scene, we created features that described the surfaces, their orientations in the frame, and relative position to the subject (Lescroart and Gallant, 2019). These features used the depth frames computed for the depth and spatial semantic feature spaces. We then used another post-processing effect to determine the normal vector (relative to the camera) of the surface under each pixel. We again divided the frame up into 5 horizontal bins that each span 22 degrees in the virtual environment. We then created 10 log-spaced distance bins between 0 and 1000 m from the subject. We then created 9 surface orientation bins: up, down, left, right, straight ahead, and four oblique directions. Within each combination of horizontal and distance bin, pixels were sorted into the closest of the 9 orientation bins. Additionally, we included one extra channel for the sky, which has an undefined orientation vector, for a total of 901 features.

3.2.9.9. Controls

The motor actions of the subjects are directly reflected by the corresponding control inputs. We therefore used included features for steering wheel angle, brake input, accelerator input, and gear selection. The steering wheel angle, brake input, and accelerator inputs were continuous values, while the gear selection was a binary indicator for a button press. There is a total of 4 features.

3.2.9.10. Affordances

Here, we define navigational affordances as where the subject is legally allowed to drive. As such, this feature space essentially captures the roads and parking lots around the subject. We use the demo files to parse the location and heading of the subject's car at each frame. We then build a series of log-polar spatial bins around the subjects' position. There are 16 directions, and in each direction, there are 10 log-spaced distance bins from 0 to 100 m away from the subject. The bins are rotated such that a reference direction is always aligned with the subject's heading at each frame. We term this bin layout "egocentric log-polar bins" to reflect that they are referenced to the immediate heading of the subject, that they are in polar coordinates, and that the distance bins are log-spaced.

We constructed a map of the extent of all the roads and parking lots in the map. Each bin is the compared against this map, and the fraction of each bin that overlaps with this navigable space is recorded as the feature value. Unlike the previous feature spaces, the affordance features are not computed from the video frames, despite the semantic segmentation providing a trivial method of finding roads in the video frames. We do so to capture affordances that are beyond the current visible space, as the view in the virtual world can only show what is in front of the subject. There is a total of 160 features.

3.2.9.11. Route progression phase (spatial)

The route progression feature spaces were inspired by a rodent study that found RSC neurons that demonstrated spatially periodic firing fields when rats ran along a track (Alexander and Nitz, 2017). Some neurons had a single firing field, while others show 2 or 4 firing fields evenly spaced along the track. These firing field periodicities were conserved across differently-shaped tracks, suggesting that they encoded generalized position along any path. This study modeled the firing fields with pairs of sinusoidal functions that swept through 1, 2, 3, 4, 6, and 8 cycles along the length of the track. The sinusoids are offset in phase by 90 degrees, and form quadrature filters that can capture periodic activity without any assumptions about phase.

Here, we adapted these features to the paths that the subjects took in the virtual world. On each trial, we linearized the path that the subject took. Each route was then normalized to a length of 1. Then, we built sine and cosine functions that swept through 1, 2, 4, and 8 cycles along this linearized path. These features reflect an encoding of route progression in the Fourier domain. At each frame, we used the position of the subject along the path to compute the value of these functions for a total of 8 features.

3.2.9.12. Route progression phase (temporal)

The temporal route progression feature space is closely related to the spatial route progression feature space. While the spatial route progression features were based on the distance elapsed, the temporal progression features are based on time elapsed. For each navigational segment, we determined the total time that the subject took. We then built sine and cosine functions that swept through 1, 2, 4, and 8 cycles in this amount of time. At each frame, we used the time elapsed since the beginning of the segment to compute the value of these functions for a total of 8 features. This feature space differs from the spatial progression features in that it is not speed-dependent. For example, when the subject is waiting at a red light, the spatial route progression features remain constant, while the temporal progression features continue to change.

3.2.9.13. Binned route progression (spatial)

The Fourier encoding of route progression implicitly assumes that neurons display periodic tuning to progression along the route. In (Alexander and Nitz, 2017), some cells were shown be partially periodic, e.g. they displayed spatially periodic firing fields in one half of the route. Such tuning profiles could not be easily captured by our Fourier basis features in a linear model, as they require interactions between features, or an exceeding large number of features that is biologically implausible. Thus, we created a complementary set of features, which divide the route into 1, 2, 4, 8, 12, and 16 bins. The 1-bin feature serves as an indicator feature that subject is actively navigating to a destination, rather than waiting during the intertrial delay period. At each frame, we determined which spatial portion of the route that the subject is in, and set the corresponding feature values to 1 and others to 0 in each set of bins. There is a total of 43 features.

3.2.9.14. Binned route progression (temporal)

We applied the same rationale in section 3.2.9.12 to the spatial binned route progression feature space. We created a corresponding set of binned temporal progression features that divide the route into 1, 2, 4, 8, 12, and 16 bins. At each frame, we determined which temporal portion of the route that the subject is in, and set the corresponding feature values to 1 and others to 0 in each set of bins. There is a total of 43 features.

3.2.9.15. Turns (spatial)

A route can be divided into smaller linear segments joined by turns. Rodent neurophysiology has shown that PPC neurons are selective for these individual segments (Nitz, 2006). To capture any such representations in our experiment, we adapted the route progression features to these smaller segments. We define "linear segments" as segments between turns. Thus, driving along a curving road is considered a linear segment, and continuing straight through an intersection is considered to be within a linear segment. Each navigational segment was manually broken up into linear segments, separated by turns. We then built sine and cosine functions that swept through 1, 2, 4, and 8 cycles along the length of each linear segment. At each frame, the position of the subject along the linear segment was used to compute the value of these functions for a total of 8 features.

3.2.9.16. Turns (temporal)

As with the route progression features, we also built a temporal version of the spatial turn progression features. We determined the time it took subjects to drive through each linear segment, and built since and cosine functions that swept through 1, 2, 4, and 8 cycles in these amounts of time. At each frame, the time elapsed since the beginning of the current linear segment was used to compute the value of these functions for a total of 8 features.

3.2.9.17. Destination vector

Previous studies have shown that the human brain tracks the proximity to a goal (Brown et al., 2016). We therefore built a feature space to encapsulate an egocentric vector to the destination. At every frame, we used the demo files to extract the position and heading of the subject, and also the location of the destination. From these values, we determine an egocentric vector from the subject to the destination in polar coordinates. We then took the log transform of the distance component of this vector. In polar egocentric space, we constructed a set of 2D Gaussian functions. In each of 16 equally-spaced radial directions, we placed the center of 12 equally-spaced 2D Gaussians functions between 1 and 12 log units from the subject. The standard deviation of these Gaussian functions was set to be 1/6 that of the distance between neighboring bins. The value of these Gaussian functions at the log-transformed destination vector were used as features. We term this process soft-binning the values. Features were zeroed during the periods between successive navigational segments. We include an additional indicator variable that is 0 when the subject is navigating, and 1 during the intertrial delay period. There is a total of 97 features.

3.2.9.18. Destination-anchored vector

The destination vector is an egocentric vector anchored to the current position and heading of the subject. An equally valid description of the relationship between the subject and destination could be anchored to the destination. We therefore also took the log-distance allocentric vector from the destination to the position of the subject, and soft-binned this vector at every frame into another set of 97 features. Note that these features are not equivalent to the destination vector. Whereas destination vector features overrepresent the space around the subject, the destination-anchored vector overrepresents the space around the destination.

3.2.9.19. Grid cells

Rodent electrophysiology studies showed that cells in the entorhinal cortex metrically measure space using a hexagonal lattice basis (Fyhn et al., 2004; Sargolini et al., 2006). We therefore developed a set of features that reflect this basis. We used hexagonal lattices with spacings of 480 m, 240 m, 120 m, 80 m, 40 m, 20, and 10 m in physical space. For each lattice, the lattice was constructed by placing 2D Gaussian functions at the vertices. These Gaussian functions had a standard deviation of 1/12 the spacing of the lattice to account for the phase-offset copies of the lattice. Each lattice forms a basis function. For each spacing, we included variants with spatial phase offsets of 1/3 and 2/3 the spacing, and also rotated copies at 20 and 40 degrees of rotation. The value of each basis function at each frame was computed by taking the maximum value of these associated 2D Gaussian functions evaluated at the position of the subject. There is a total of 189 features that tiled the map.

3.2.9.20. Destination grid representation

Place cell spatial firing fields can be derived from multiple grid cell inputs (Solstad et al., 2006). Thus, at any moment, the subject's position can be encoded using only the grid cell features. Because the destination is of particular interest during navigation, we postulated that its position may also be represented. Therefore, we created a second set of grid cell features using the same parameters described above. During each trial, these features were used to represent the position of the destination. During the intertrial delay, these features were zeroed. There is a total of 189 features.

3.2.9.21. Beeline distance elapsed

Path integration is commonly assumed to be computed with respect to some starting position. For each navigation segment, we defined the starting position as the position of the subject when the destination was cued. At each frame, we computed the beeline distance from the current position of the subject to the starting position. The log transform of this distance is then soft-binned into 12 bins equally spaced between 1 and 12 log units for a total of 12 features. Like the destination vector features, the feature values were zeroed in the periods between successive navigational segments.

3.2.9.22. Beeline distance remaining

At every frame, we also computed the beeline distance between the subject and the destination. Like the beeline distance elapsed features, the log transform of the beeline distance remaining was soft-binned into 12 bins equally spaced between 1 and 12 log units for a total of 12 features. Note that the beeline distance remaining is implicitly captured in the destination vector feature spaces. However, that does not preclude a separate representation of only the distance. Additionally, in other experiments, we had found that explicit features often explained further variance even if they are implicitly captured by other feature spaces.

3.2.9.23. Path distance elapsed

Very rarely do subjects travel in a straight line between two points. It is thus likely that, in addition to the beeline distance, the brain also tracks the actual distance traveled. At every frame, we computed the actual distance traveled from the starting position. The log transform of this distance is then soft-binned into 12 bins equally spaced between 1 and 12 log units for a total of 12 features. Note that unlike the route progression features, the path distance elapsed features are not normalized, and metrically reflect the distance traveled.

3.2.9.24. Path distance remaining

Likewise, we also computed the path distance remaining between the subject and the destination. At every frame, we used the demo files to look ahead in time, and calculate the distance that subject will travel from their current position to the destination. The log transform of this distance is also soft-binned into 12 bins equally spaced between 1 and 12 log units for a total of 12 features. Like the path distance elapsed features, these features are also not normalized and metrically reflect the distance remaining.

3.2.9.25. Egocentric path integration

The scalar beeline distance remaining feature space is complemented by the vector destination vector feature space. The egocentric path integration provides a similar vector complement to the beeline distance elapsed feature space. At every frame, we compute the polar egocentric vector from the subject's position to the start point. The distance component is then log-transformed, and the resulting vector is soft-binned into 8 directions with 12 distances per direction. We also include an indicator variable that is 0 when the subject is navigating, and 1 during the intertrial delay. There is a total of 97 features.

3.2.9.26. Allocentric path integration

The egocentric path integration feature space is referenced to the current heading of the subject. We also developed an allocentric version of these features in which the vector is referenced to the global north. The distance component is also log-transformed, and the resulting vector is also soft-binned into 8 directions with 12 distances per direction. Again, we include an indicator variable that is 0 when the subject is navigating, and 1 during the intertrial delay. There is a total of 97 features.

3.2.9.27. Vehicles

Navigation in the real world is a dynamic process that may require interaction with, and therefore representation of, other agents. Indeed, electrophysiology studies show that both rat and bat hippocampus contains social place cells that represent the position of other animals in the environment (Danjo et al., 2018; Omer et al., 2018). It is likely that the human brain contains similar representations. In the virtual world used in this experiment, other vehicles may influence the immediate navigational decisions of the subject. We developed a set of features to capture these possible representations.

At every frame, we used the demo files to extract the position and view vector of the subject, and also the position of all other vehicles in the world. Next, we constructed a set of egocentric log-polar bins around the subject. For each of 16 directions, we placed 10 log-spaced distance bins from 5 m to 200 m away from the subject. We then counted the number of vehicles in each bin. Bin values were then used to create a total of 160 features.

3.2.9.28. Pedestrians

In addition to vehicles, the virtual world also contains pedestrians that may also affect the immediate navigational decisions of the subject. Because pedestrians differ from vehicles in both appearance and behavior, they are likely to be represented separately. Therefore, we created a separate set of features for pedestrians following the same methods for vehicles. There is a total of 160 features.

3.2.9.29. Future Path

A plan is needed for successful navigation. We operationally divide the plan into two scales: the long-term abstract plan and the short-term concrete plan. The future path feature space captures the short-term concrete plan. This feature space metrically represents where the subject plans to move in the visible world and the immediately adjacent non-visible space. We again make use of egocentric polar bins. In each of 16 directions, we place 10 bins log-spaced from 2.5 m to 100 m away from the subject. At each frame, we use the demo files to look ahead in time for the locations of the subject in future frames. We search forwards in time to either when the subject is beyond the furthest edge of the bins (100 meters from the current location), or until the end of the current navigational segment, whichever is earlier. Then, for each bin that the subject will be in in the future, we set its value to 1. We set all other bins values to 0. These 160 features capture the near-term navigational plan of where the subject plans to go.

3.2.9.30. Road graph

To complement the short-term concrete plan captured by the future path feature space, we developed the road graph feature space to capture the long-term abstract plan. Human behavioral studies suggest that larger spaces are represented on a relational basis that conserves their topologic organization, but not necessarily their metric spatial organization (Chrastil and Warren, 2014; Warren et al., 2017). Thus, we built a graphical model of the road network in the virtual city. Intersections were used as vertices and road segments between

intersections were used as edges for a total of 232 vertices and 294 edges. Furthermore, graphical planning may be a multiscale process, in which plans are conceptualized on both general (e.g., going to the grocery store) and specific (e.g., the particular route to take) levels. To capture this multiscale planning, we performed graph simplification on this road graph. In this process, the closest pair of vertices is merged into a single vertex, and the edge connecting the pair is deleted. All edges connecting to either of the two original vertices are replaced with new edges that connect to the new vertex. This step is iteratively applied to the graph until there remains only a single vertex. These new vertices and edges increasingly reflect the implicit connectivity between different parts of the map, rather than the direct road graph. Including the aggregate vertices and edges, there are a total of 463 vertices and 1042 edges. We use edges as features for a total of 1042 features.

At each frame, we determine the road edges that correspond to the path that the subject will take to the destination. We set the feature value corresponding to each of these edges to 1. These edges reflect the specific plan. Then, for the vertices that these edges connected, we find all aggregate vertices into which they collapse. We then find all edges that connect all these aggregate vertices, and also set their feature values to 1. These aggregate edge features reflect the more general plan. Taken together, these features reflect the subjects' multiscale abstract navigation plan on the road network.

3.2.9.31. Head direction

While grid cells are well-established in electrophysiology, they remain more elusive in fMRI. (Doeller et al., 2010) demonstrated evidence for a grid-like representation indirectly using fMRI. This study is premised on the basis that the axes of individual grid cells are spatially aligned. Thus, traveling in a direction aligned with the hexagonal axes should cause grid cells to fire more than if the directly of travel is misaligned. Voxels contain grid cells should therefore display a six-fold symmetry in activity. (Doeller et al., 2010) indeed show such six-fold symmetry and thus suggest that it provide evidence for grid cells in fMRI.

We developed the head direction feature space to independently test this hypothesis. We constructed sine and cosine functions that sweep through 6 cycles across 360 degrees. We used these functions instead of binning by direction, because a pair of sinusoid functions are agnostic phase offsets. Bins that are out of phase with the actual selectivity would perform poorly, whereas these functions would be able to capture the phase of the preferred direction. At every frame, we took the heading of the subject, and calculated the value of the sine and cosine functions at this direction for a total of 2 features.

3.2.9.32. Gaze direction

We defined the head direction as the allocentric heading of the subject's car in the virtual world. Because subjects are not fixating, they may not necessarily look be looking in the direction of travel. The virtual camera has a horizontal field of view of 110°. Therefore, the allocentric gaze direction may differ from the heading by up to 55°, a significant divergence.

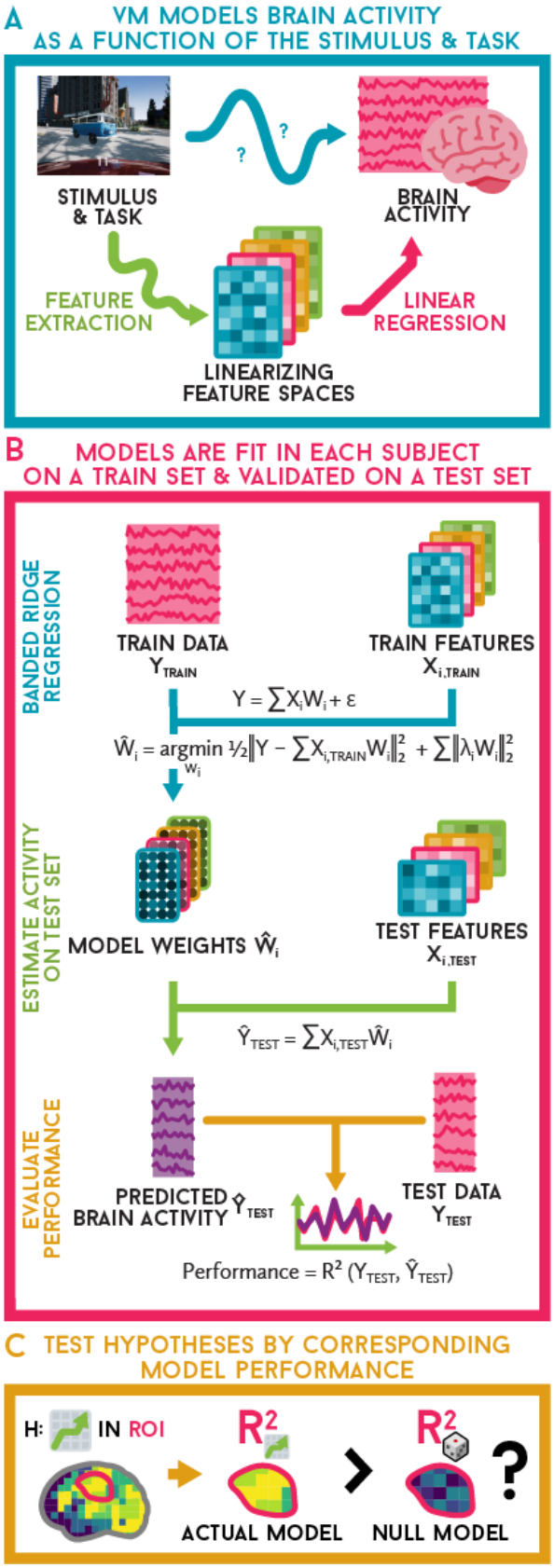


Figure 3.2 Voxelwise modeling (VM) procedure. A) Conceptually, VM treats the activity of each voxel as some nonlinear function of the stimulus and task parameters. To make models tractable and interpretable, VM nonlinearly extracts features from the stimulus and task. Brain activity is then modeled as a linear combination of these features. B) Banded ridge regression is first used to fit models simultaneously to many feature spaces in a training set. Model performances are then evaluated by predicting brain activity on a hold-out validation dataset not used in training. Doing so minimizes overfitting and ensures that models generalize beyond the dataset to which they are fit. C) The fit models can then be used to test hypotheses about representation. The significance of model performances reveals the representation of features across the brain, and model weights reveal feature tuning.

We also build features to account for this possible separate representation. At every frame, we compute the value of the same sinusoid functions used in the head direction feature space, for a total of two features.

3.2.9.33. Gaze grid

Non-human primate electrophysiology suggests that the gaze position in a head-centered reference frame is also represented on a hexagonal lattice basis. There has also been weak fMRI evidence for such representations in the human brain (Nau et al., 2018). We therefore built a set of features to represent gaze location using a grid cell-like hexagonal basis. We used hexagonal lattices with spacings of 16, 32, 64, 128, 256, and 512 pixels in screen space. For each lattice, the lattice was constructed by placing 2D Gaussian functions at the vertices. These Gaussian functions had a standard deviation of 1/12 the spacing of the lattice to account for the phase-offset copies of the lattice. Each lattice forms a basis function. For each spacing, we included variants with spatial phase offsets of 1/3 and 2/3 the spacing, and also rotated copies at 20 and 40 degrees of rotation. At each frame, the value of each basis function was evaluated at the gaze position in screen space for a total of 162 features.

Together, these 33 feature spaces encompass 21,283 features. All feature timecourses were first computed on a frame-by-frame basis, and then downsampled to the fMRI sampling rate.

3.2.10. Voxelwise modeling with banded ridge regression

We used voxelwise modeling (VM) (Deniz et al., 2019; Huth et al., 2012, 2016; Lescroart and Gallant, 2019; Naselaris et al., 2011; Nishimoto et al., 2011; Nunez-Elizalde et al., 2019) to recover maps of how navigational features are represented across the cortex (Figure 3.2). VM treats brain activity as a nonlinear function of stimulus and task parameters. Because nonlinear models are computationally intractable to fit, and are difficult to interpret, VM makes use of linearizing feature spaces. Each feature space is a nonlinear transformation of the stimulus or task, and is designed to be interpretable. BOLD activity $\boldsymbol{Y}$ is then modeled as a linear combination of features $\boldsymbol{X}$ with Gaussian noise $\boldsymbol{\varepsilon}$. This model is expressed as $\boldsymbol{Y}=\boldsymbol{XW}+\boldsymbol{\varepsilon}$, in which $\boldsymbol{W}$ are weights that map feature values on to brain activity. These weights describe how each feature drives activity in each voxel.

Because features often outnumber the number of TRs, regularized linear regression is used to solve for the optimal weights. The most common form of regularized regression, ridge regression, imposes a spherical multivariate normal prior on feature weights. In this case, the optimal solution expressed as

$$\hat{W} = \underset{W}{\operatorname{argmin}} \tfrac{1}{2}\left\|Y - XW\right\|_2^2 + \left\|\lambda W\right\|_2^2$$

in which $\boldsymbol{\lambda}$ is the regularization parameter. Intuitively, ridge regression treats all features equally.

While ridge regression is well-suited for VM with a single feature space, it is suboptimal when many feature spaces are used. A single global regularizer biases ridge regression to favor larger feature spaces at the expense of smaller ones. Each feature space may contribute differently to the activity in each voxel, and these contributions may not be correlate with

feature space size. We therefore used banded ridge regression (Nunez-Elizalde et al., 2019), which places a non-spherical multivariate normal prior on feature weights. This non-spherical prior is implemented by regularizing each feature space independently instead of applying a single global regularizer to all weights. Independent regularization treats all feature spaces with equal importance regardless of feature space size. The optimal solution to banded ridge regression is expressed as

$$\hat{W}_i = \underset{W_i}{\operatorname{argmin}} \tfrac{1}{2}\left\|Y - \sum X_i W_i\right\|_2^2 + \sum \left\|\lambda_i W_i\right\|_2^2$$

in which $\boldsymbol{i}$ is the index of feature spaces. Note that if $\boldsymbol{\lambda}_i$ are the same for all feature spaces, then banded ridge regression reduces to regular ridge regression. At the same time, we model the hemodynamic response function with a finite impulse response (FIR) filter separately for each feature in each voxel. The FIR filter is implemented by including copies of feature timecourses at 5 temporal delays corresponding to 0, 1, 2, 3, and 4 TRs.

For n models with k possible regularizers, the size of the regularization hyperparameter space scales at a rate of $O(k^n)$. It is computationally infeasible to fully explore the high-dimensional hyperparameter space associated with 30+ models. In previous experiments, we had used a tree-structured Parzen (TPE) search performed on CPU (Bergstra et al., 2013). We had found that TPE empirically becomes increasingly inefficient with higher numbers of models, and is prone to become trapped in local minima. Therefore, in this experiment, we instead used a more efficient GPU implementation of the banded ridge algorithm (Dupre la Tour et al., n.d.). This implementation uses a random Dirichlet simplex sampling in the hyperparameter space. We selected concentration parameters such that the sampling is biased towards the vertices on the simplex. In other words, we bias the hyperparameter choices such that, for each voxel, most feature spaces are heavily regularized. This reflects an implicit sparsity prior that each voxel likely only represents a few of all possible feature spaces, i.e., that the cortex is specialized.

The optimal regularizer per feature space per voxel, and the shape of the FIR filter, were estimated empirically by cross-validation. We used the coefficient of determination (R^2) as the performance metric. We randomly sampled 40,000 different combinations of hyperparameters, and performed 10-fold cross-validation twice to determine performance. The optimal combination of hyperparameters was selected independently for each voxel. After models were fit, they were evaluated by predicting brain activity in a hold-out dataset not used in model estimation. Again, we used R^2 as the metric for performance. We divided the overall model, which includes all W_i corresponding to all feature spaces, into individual models, each of which corresponds to a single W_i for one feature space. For each voxel, we computed both overall model performance, and also performance for each individual model.

3.2.11. Mapping navigational representations across the cortex

The VM procedure produces high-dimensional models that contain over 21,094 features distributed across over 33 feature spaces. In previous studies, we had examined the representation of features within a single feature space across the cortex via dimensionality

reduction on the feature weights (Deniz et al., 2019; Huth et al., 2012, 2016). Here, we extend the same method to examine the representation of multiple feature spaces across the cortex. In each voxel, individual model prediction performances form a 33-dimensional vector that describes which feature spaces are represented. We applied PCA to these model performance vectors to discover the most important dimensions in navigational representations across the cortex.

3.2.12. Cross-subject comparisons and group-level analyses

These previous analyses are performed in the native space of each subject. In order to better understand individual differences in navigational representations, we also mapped representations at the group level. Because all subjects were fit with the same feature spaces, the individual model performance space was used for group analyses. Doing so allows us to avoid averaging across subjects and retain the most information for modeling.

To reduce noise, we thresholded voxels by overall model prediction performance. In each subject, we selected voxels in which the overall model explained at least 1% of the variance, and performed PCA on their model performance vectors. We then projected each subject into the common space formed by these PCs. To examine how well these PCs generalized across subjects, we also performed a leave-one-subject out analysis. For each subject, we calculated PCs using the best voxels from the other two subjects. That subject was then projected into this group space.

These PCs describe the distribution of navigational representations across the cortex. To test their significance, we use the principal components of feature space powers as a null distribution. We define the feature space power as the norm of all features within a feature space at each TR. A 1-dimensional power timecourse is thus computed for each feature space. Because larger feature spaces would have larger norms, these timecourses were z-scored. This normalization equalizes feature spaces to reflect their equal weighting during the model fitting process. Because each subject saw different stimuli, feature space power was computed for each subject with their own data. Feature space power was then bootstrapped to form a null distribution for the PC space.

This provides a valid comparison for the distribution of navigational representations across the cortex. If our model did not capture any significant distributions of cortical navigational representations, then any variance in these model performance vectors would reflect the variance in feature space power. On the other hand, if the models are instead capturing variance in the brain's representation of navigational information, then the model performance PC space would differ significantly from the feature power PC space.

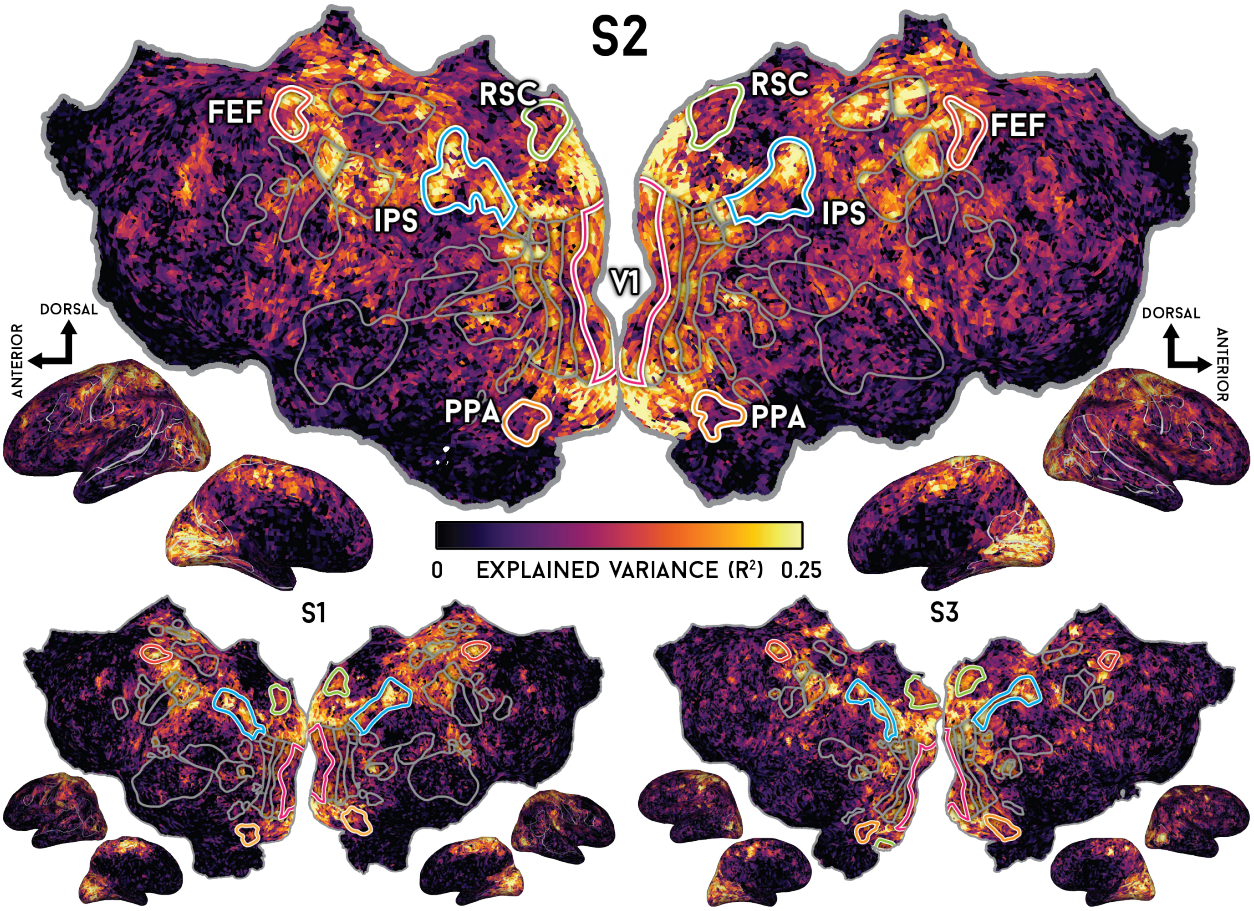


Figure 3.3 Overall prediction performance. BOLD activity in each voxel in each subject was modeled as a combination of 21,283 features across 33 feature spaces. Banded ridge regression was used to simultaneously fit all 33 models and find the optimal combination of model parameters. Models were then evaluated by predicting activity on a hold-out validation dataset not used during model training. Performance is measured by the amount of variance that models are able to explain in this validation set. We plot the overall aggregate performance of all 33 models on the flattened cortex for each subject, and inflated cortices are shown for reference. Models capture robust activity across the cortex, include visual, somatosensory, motor, parietal, and prefrontal cortex. We find similar patterns of activity across the cortex in all subjects.

3.3. Results

To better understand how the human brain represents navigational information, we developed an active navigation experiment in fMRI. In this experiment, subjects used an MR-compatible steering wheel and pedal set to drive a virtual in a large virtual city. Prior to scanning, subjects learned the layout of the city. Subjects performed a taxi driver task while BOLD activity was recorded. We then applied voxelwise modeling to the brain activity timecourses. Banded ridge regression was used to simultaneously fit models to 33 feature spaces that captured a variety of navigational information. Model performances were then evaluated by predicting activity timecourses on a hold-out validation dataset. For each voxel, these model performances form a 33-dimensional vector describing the amount of variance each model can explain.

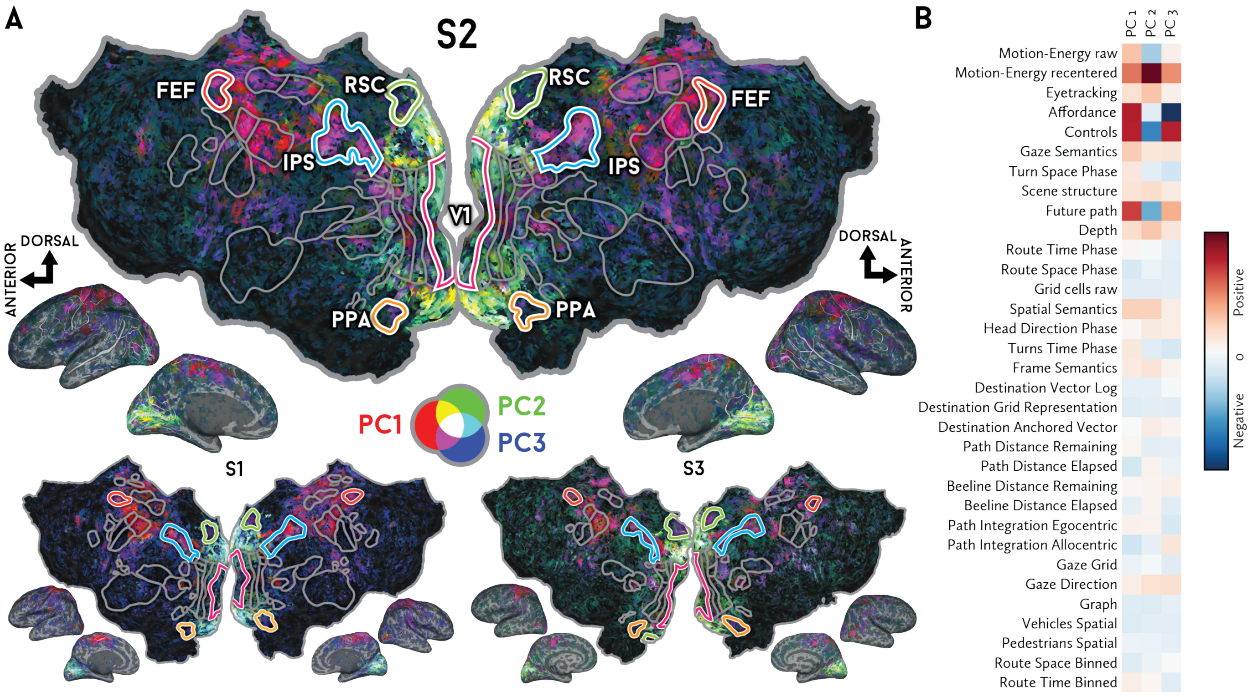


Figure 3.4 Principal components of navigational representations are consistent across subjects. To examine what features spaces are represented by different brain regions, we computed the amount of variance each of the 33 models is able to explain in each voxel. Thus, each voxel is described by a 33-dimensional vector that captures individual model performances. We then selected all voxels in which these models were able to explain at least 1% of the variance. We then performed principal components analysis (PCA) on the individual model performance vectors for these voxels to discover the most important dimensions of navigational representation across the cortex. A) We map the first three PCs to the red, green, and blue channels, respectively, and plot them on the cortical surface. Based on the cortical areas to which they are mapped, we term the first PC to be "motor," the second PC to be "visual," and the third to be "navigational." B) PC weights across feature spaces for the top three PCs. Each column corresponds to one PC, and each row is the weight of each feature space on to the top three PCs. The first PC show high weights for affordance, controls, and future path. The second PC shows high weights for motion-energy. The third PC shows high weights for controls and future path. These top three PCs are consistent across subjects, indicating that representations of navigational information during active navigation is largely consistent across subjects.

3.3.1. Subjects are able to successfully navigate in the virtual world

All subjects successfully learned the world and were able to navigate in the virtual world. We defined a *route optimality* metric to quantify the navigational abilities of subjects. Route optimality is defined as the ratio between the length of the subject's path to the distance of a path found by A* pathfinding. We computed the route optimality for all routes that the subjects took during data collection. During data collection, the median route optimality was 1.05 for S1, 1.01 for S2, and 1.23 for S3.

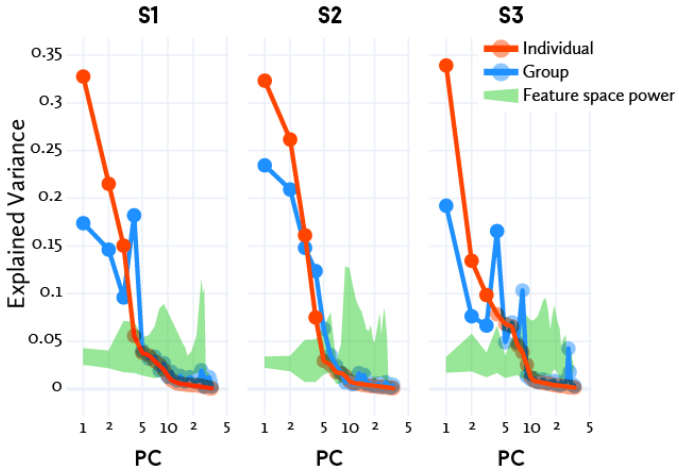


Figure 3.5 PCA was applied to 33-dimensional individual model performance vectors to discover the most important dimensions in navigational representations across the cortex. Here, we show the amount of variance in navigational representations across the cortex explained by these PCs. Red lines show the amount of variance explained by each subject's own PCs. Blue lines show the amount of variance explained by group PCs calculated from the other two subjects. We compared these against PCs computed from the timecourses of feature space power for each subject. Shaded green regions show 95% confidence intervals on feature space power PCs. Confidence intervals on group and individual PCs are too small to be shown. Solid marks show PCs that explain significantly more variance than feature space power PCs ($P < 1e\text{-}3$, bootstrap test). Four group PCs consistently explain more variance in all subjects, suggesting that there may be four distinct functional networks that support navigation.

3.3.2. Active navigation elicits rich activity across the cortex

In all subjects, the voxelwise models explain variance across much of the cortex. In Figure 3.3 we show the overall model performance on the cortex in all subjects (for individual model performances, see Supplementary Figures 1-6). These areas include many known functional ROIs, including V1-V4, IPS, MT, RSC, OPA, PPA, FFA, EBA, M1, SMA, and FEF. Additionally, they include areas outside known ROIs, include the precuneus, parts of the posterior parietal cortex, dorsolateral prefrontal cortex, and medial prefrontal cortex. These models are able to explain at least 1% of the variance (corresponding to approximately at least 0.1 correlation) in 29,370 ± 3,899 (mean ± std) voxels in each subject (35.6% ± 5.8% of cortical voxels). These results show that active navigation engages multiple distributed networks in the brain, and that these networks are consistently engaged between multiple subjects.

3.3.3. Voxelwise models of active navigation reveals multiple networks of brain regions

To understand the most important dimensions of navigational representations across the cortex, we applied PCA to the 33-dimensional model performance vectors (Figure 3.4). We inspected these PCs by their projections on the cortex, and also by the loading of the different models on to the PCs. We find that the first three PCs could be termed the "motor," "visual," and "navigational" networks. The visual and motor networks are largely confined to networks of known ROIs. The visual network spans much of the visual cortex, including anterior visual

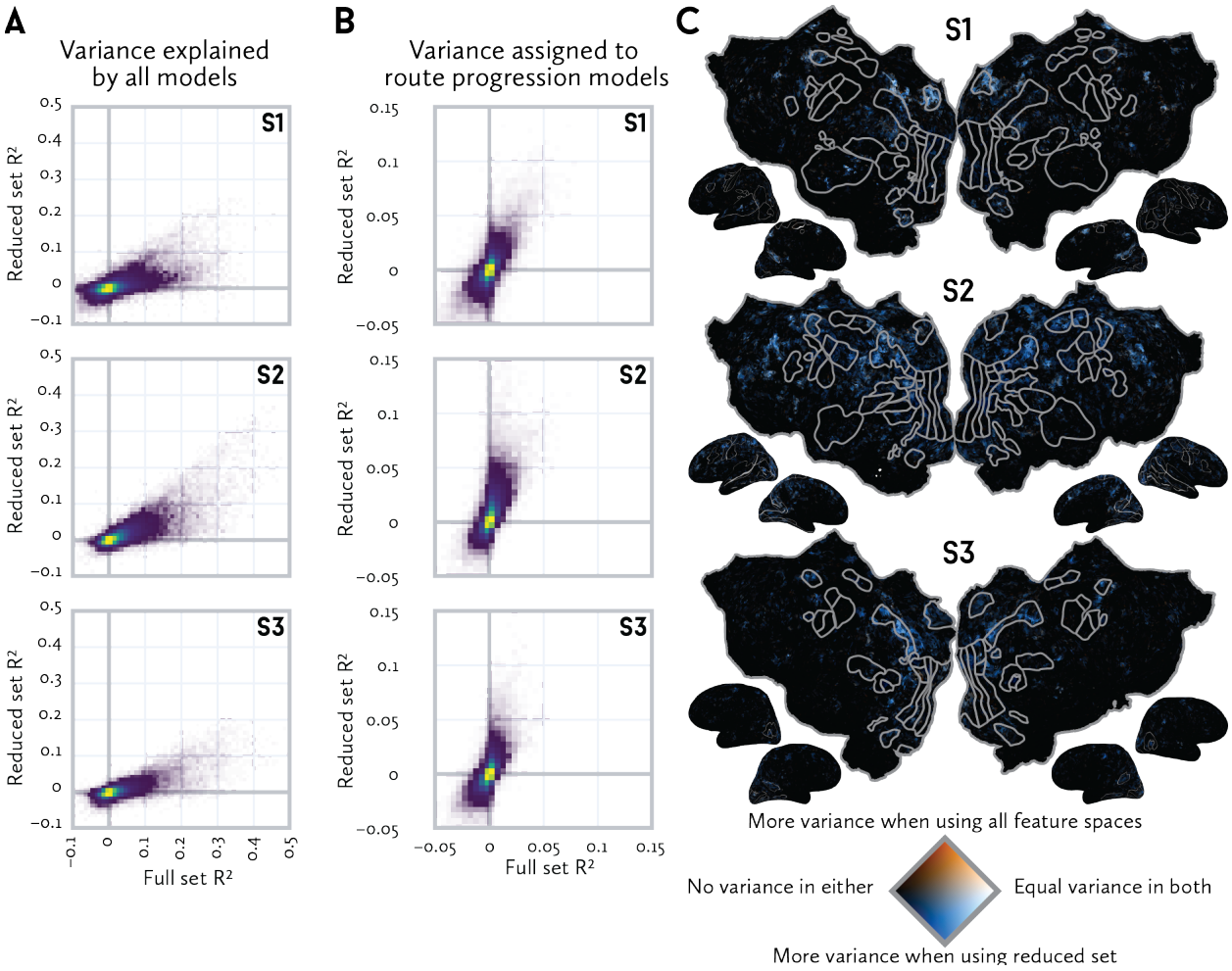


Figure 3.6 Correlations between features can inflate model performances. Banded ridge can mitigate this issue, but an appropriate set of feature spaces must be included in the model fitting process. Here we re-fit models using only the retinotopic motion-energy and the four route progression feature spaces. A) 2D histograms of overall prediction performance in all subjects comparing models fit with all 33 feature spaces (full set) versus with only five feature spaces (reduced set). The vertical axis is the distribution in the reduced set, and the horizontal axis is the distribution in the full set. Models from the reduced set consistently explained less variance across the cortex in all subjects. B) However, the opposite is true for the route progression models. Here we plot 2D histograms of the amount of variance explained by the route progression models were able to explain. In the reduced set, the route progression models consistently explained more variance than they do in the full set. C) We compare route progression model performances in the full and reduced fits across the cortex. The route progression models explain no variance in black voxels, more variance in the reduced set in blue voxels, more variance in the full set in orange voxels, and equal variance in both condition in white voxels. Extraneous variance is attributed to the route progression models in the reduced set fit due to stimulus correlations. To minimize biases, it is imperative that all possibly relevant feature spaces are used when fitting models to data from naturalistic experiments. To our knowledge, the 33 feature spaces here are the most comprehensive set of features used in a single experiment to date, and provides models with the least amount of bias currently possible under naturalistic conditions.

areas, and also the intraparietal sulcus. The motor network spans the somatosensory and motor cortices and the supplemental and pre-motor areas. Unlike the other two networks, the navigation network is not confined known functional ROIs. The navigation network includes RSC, OPA, and PPA, but also regions in precuneus, parietal cortex, and also both the dorsolateral and ventral prefrontal cortices. In the parietal cortex and precuneus, the navigation network immediately borders the IPS, the somatosensory cortex, and anterior visual regions. These PCs show that active navigation elicits much activity beyond known ROIs. This activity may be overlooked by more traditional ROI-based approaches, or may not be elicited by less naturalistic tasks.

We then used a leave-one-subject-out approach to examine the consistency of these group-level networks across subjects (Figure 3.5). For each subject, we constructed PCs using the model performance vectors of all other subjects. We then projected that subject's model performance vectors on to these PCs, and examined how much variance these PCs were able to explain. We compared these against the group PCs against each subject's own PCs. To test that these PCs reflect the brain's navigational networks and are not merely stimulus-driven, we used the power of the feature spaces as controls. Four group PCs consistently explain more variance in all subjects, suggesting that there is a four-dimensional distribution of navigational representation that is shared across individuals. Note that in this section, we performed PCA on the individual model performance vectors. This analysis reveals where different features are represented, and the similarities of these cortical locations across subjects. This analysis does not reveal the cortical tuning for individual features, or how they may differ across subjects. We examine cortical tunings in other, targeted analyses of individual models.

3.3.4. Extraneous variance is assigned to models when few feature spaces are used

In this study, we simultaneously fit 33 models to capture a variety of navigationally-relevant information. In complex interactive tasks such as driving, many features are necessarily correlated. Banded ridge can decorrelate models and assign variance explained to each model. Implicitly, however, this process requires the appropriate feature spaces to be included during the model fitting process. When insufficient feature spaces are used, variance could be mis-attributed to models.

As a case study, we examined how model performances changed when only a subset of feature spaces is used to fit models (Figure 3.6). In a reduced set, we fit models using only low-level vision (retinotopic-motion energy) and route progression (spatial route progression and temporal route progression, in both their phase and binned variants) feature spaces. In all subjects, we find that much more variance is attributed to the route progression models in this reduced set than when all feature spaces are used. In the reduced set, the route progression models explained at least 1% of the variance in 10,675 ± 3659 voxels (mean ± std across subjects), as opposed to 3040 ± 842 voxels when all feature spaces were used ($p = 0.045$, t-test). In these voxels, the route progression models explain an average of 2.50 ± 1.70% (mean ± std across all subjects) of the variance (max 15.26 ± 1.78 % mean ± std across subjects) in

the reduced set, and 1.73 ± 0.80% variance (max 6.66 ± 1.12 %) in the full set (p = 0.0, t-test). Many of the voxels explained by the route progression models in the reduced set were also explained by other navigational feature spaces in the full set.

The retinotopic motion-energy model explained at least 1% of the variance in variance in 9,592 ± 1,646 voxels (mean ± std across subjects), and 5,710 ± 1,564 voxels in the full set. In these voxels, the route retinotopic motion-energy model explains an average of 3.30 ± 3.30 % of the variance (max 25.70 ± 6.44 % mean ± std across subjects) in the reduced set and 3.53 ± 3.42 % of the variance (max 23.30 ± 5.43 %) in the full set. Many of the voxels explained by the retinotopic motion-energy model in the reduced set were also explained by motor controls and related models.

These results show that, in the absence of more appropriate feature spaces more correctly capture representations, extraneous variance will be assigned to models because of stimulus correlations. Models fit with suboptimal sets of feature spaces will therefore show inflated model performances and misattribute variance to feature spaces. This misattribution of variance can lead to spurious conclusions about representation when fewer feature spaces are used to fit models. In more extreme cases, they may lead to completely erroneous conclusions about representation. This result highlights the pitfalls of not fully modeling brain activity, particularly in naturalistic settings.

3.3.5. Voxelwise models are used to test hypotheses in a naturalistic setting

Many existing hypotheses on how the human brain represents navigational information are derived from two categories of experiments. The first are human neuroimaging experiments that used tightly controlled and unnatural stimuli. The second are non-human experiments that used model organisms such as rodents. In both cases, it is crucial to examine whether these results generalize to human brains operating under more naturalistic conditions.

Here, we use the voxelwise models derived from this experiment to test many hypotheses about navigational representations. Each hypothesis is formulated in terms of either model performance or model weights in particular ROIs. Here, we test five diverse hypotheses about navigational representations.

3.3.5.1. Hypothesis: RSC and PPC track position along a route

In rodents, RSC and PPC were found to represent information related to the route that the animal traversed. When rats ran along a linear track, PPC neurons displayed spatial firing fields anchored to different portions of routes (Nitz, 2006). When rats ran along a closed circuit, RSC neurons were found to display spatially periodic firing fields along the circuit (Alexander and Nitz, 2017). These representations were invariant to the shape of the tracks, suggesting that these neurons abstracted represented the route. We asked whether this generalizes to the human brain (Figure 3.7).

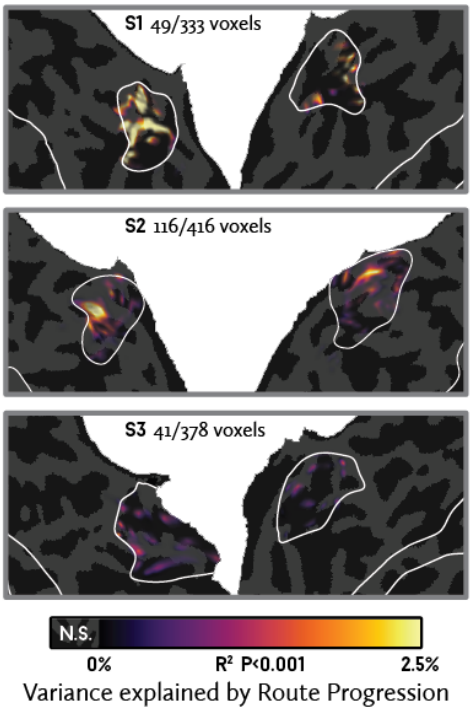


Figure 3.7 Voxelwise models of navigational representations in the human brain can be used for hypothesis testing. Rodent electrophysiology (Alexander & Nitz 2017) suggests that RSC represents the position of an animal along a complex track. To test this hypothesis in the human brain, we examined the amount of variance the route progression models are able to explain in RSC. We find that 17.8 ± 7.3 % (mean ± std.ev. across subjects) of RSC voxels represented route progression information ($p < 1e\text{-}3$, permutation test). In these voxels, the route progression models explained 2.0 ± 1.2% (mean ± std.ev. across all subjects and voxels, max 5.1 ± 1.9 % across subjects) of variance. We show these results on the cortical surface in RSC in each subject. Voxels are colored by the amount of variance the route progression models are able to explain. Transparent voxels do not significantly represent this information. These results show that human RSC does indeed represent the position of the subject along a route. However, this representation is represented in only a minority of voxels in human RSC, and do not dominate the activity of these voxels.

To test this hypothesis, we selected voxels from the RSC and IPS in our subjects. We used IPS as the most homologous region to rodent PPC. For each voxel, we then determined the amount of variance explained by the route progression models. We find that 17.8 ± 7.3 % (mean ± std.ev. across subjects) of RSC and 9.7 ± 3.6 % of IPS voxels represented route progression information ($p < 1e\text{-}4$, permutation test). Of these voxels, the route progression models explain 2.0 ± 1.2% (mean ± std.ev. across all subjects and voxels, max 5.1 ± 1.9 % across subjects) of the variance in RSC and 1.8 ± 0.9 % (max 4.6 ± 0.7% across subjects) of variance in IPS. These results show that human RSC and IPS are likely to represent path progression information in an active navigation task. However, this information is represented by a minority of voxels withing these ROIs, and do not dominate the activity of these voxels.

As a control, we also tested voxels in the mouth region of the primary motor cortex (M1M). In these voxels, 8.6 ± 3.5 % (mean ± std.ev. across subjects) were significantly predicted ($p < 1e\text{-}4$, permutation test). Of these voxels, the route progression models explain 0.9 ± 0.5% (mean ± std.ev. across all subjects and voxels, max 2.6 ± 0.5 % across subjects) of the variance. We note that the route progression models explain much less variance in many fewer voxels in this control ROI. As noted in section 3.3.4, there may still be unmodeled stimulus correlations that cause extraneous variance to be assigned to models, and M1M is unlikely to represent route progression.

3.3.5.2. Hypothesis: RSC represents path integration information

Human neuroimaging studies suggest that RSC represents a homing vector during navigation (Chrastil et al., 2015, 2016). However, these studies were performed with subjects passively watching a video recorded in a sparse virtual environment. We tested whether this

generalizes to active navigation in a more naturalistic world. For each voxel in RSC in each subject, we determined the amount of variance explained by the path integration models. We find that 24.0 ± 12.7 % (mean ± std.ev. across subjects) of RSC voxels represents path integration information ($p < 1e\text{-}4$, permutation test). Of these voxels, the path integration models explain 1.6 ± 0.9% (mean ± std.ev. across all subjects and voxels, max 4.4 ± 0.6 % across subjects) of variance. These results suggest that path integration information is likely represented in RSC. However, like route progression, path integration representation is found in only a minority of voxels and do not dominate the activity of these voxels.

We also examined whether path integration and route progression were represented by the same voxels. In RSC, we found that 4.5 ± 3.0% of RSC voxels represented both types of information. This small overlap suggests there exist functionally separate subpopulations within RSC that represent these two types of information.

3.3.5.3. Hypothesis: PPC represents planned future movements

In non-human primates, PPC neurons represent planned actions by the animals (Andersen and Buneo, 2002; Snyder et al., 1997). These studies were not navigation tasks, but rather reaching tasks in which animals planned arm movements. However, because PPC is known to represent navigational information (Sato et al., 2006), it may also represent information about future navigational actions. Again, we used IPS as the most homologous region to non-human PPC. For each voxel, we then determined the amount of variance explained by the future path model. We find that 2.4 ± 1.9 % (mean ± std.ev. across subjects) of IPS voxels represents future path information ($p < 1e\text{-}4$, permutation test). Of these voxels, the future path model explains 1.3 ± 0.7% (mean ± std.ev. across all subjects and voxels, max 3.0 ± 1.3 % across subjects) of variance. These results suggest that future path information is at most represented only by a small fraction of IPS voxels, and do not contribute significantly to IPS activity.

3.3.5.4. Hypothesis: FFA, EBA, and TPJ represent the presence of other agents

Human neuroimaging studies have revealed that the FFA and EBA are selective for faces and body parts, respectively (Downing et al., 2001; Kanwisher et al., 1997). Additionally, the EBA is also likely selective for actions and movement. However, in these studies, subjects viewed clear images or videos of faces and bodies that are similar to close, personal settings. Humans form the dominant object in these stimuli. It is unclear whether these representations persist in less ideal conditions in which persons occupy a much smaller portion of the scene, are sometimes occluded, and do not dominate the scene.

The temporoparietal junction (TPJ) is engaged during theory of mind tasks (Saxe and Kanwisher, 2003). These studies have mostly used text-based narratives or image-based narratives. During active navigation in a dynamic environment, it is likely subjects engage in some theory-of-mind processes to attempt to estimate the intentions of pedestrians and

other vehicles. To examine whether EBA and FFA represents humans under these more dynamic conditions, and whether TPJ is engaged during this navigation task, we examined the weights for "pedestrians" and "vehicles" in the gaze semantics model in these ROIs.

We find 19.2 ± 12.1 % (mean ± std across subjects) of FFA voxels, 45.5 ± 22.9 % of EBA voxels, and 16.3 ± 11.0 % of TPJ voxels were significantly selective for pedestrians ($p < 1e\text{-}3$, permutation test). Additionally, we find that 5.7 ± 7.0 % (mean ± std across subjects) of FFA voxels, 11.3 ± 6.2 % of EBA voxels, and 10.5 ± 6.6 % of TPJ voxels were significantly selective for vehicles ($p < 1e\text{-}3$, permutation test). Furthermore, 1.3 ± 1.4 % (mean ± std across subjects) of FFA voxels, 4.3 ± 3.2 % of EBA voxels, and 2.3 ± 2.3 % of TPJ voxels were selective for both pedestrians and vehicles. These results show that humans are represented by a large fraction of FFA and EBA voxels, even under these complex dynamic conditions. Additionally, FFA and EBA also represent vehicles, suggesting that these canonically human-selective areas may be selective for other agents. Furthermore, TPJ is engaged in representing humans and vehicles, suggesting that they elicit theory-of-mind activities. However, the small overlap of human- and vehicle-selective voxels suggest that there may also be functionally distinct populations in all three ROIs.

3.3.5.5. Hypothesis: RSC represents heading direction

Human neuroimaging studies suggest that RSC represents allocentric head direction (Shine et al., 2016; Vass and Epstein, 2013). These studies used static stimuli that were not representative of active navigation in the real world. To test the hypothesis that RSC represents allocentric head direction, we used the head direction model from our data. For each RSC voxel, we determined the amount of variance explain by the head direction voxel. In one subject, the head direction model failed to explain any significant variance in RSC voxels (0.00 ± 0.12% mean ± std.ev. variance explained). In the other two subjects, only 2.15 ± 1.55 % (mean ± std.ev. across subjects) of RSC voxels represent head direction information ($p < 1e\text{-}4$, permutation test). Of these voxels, the head direction model explains 0.49 ± 0.24% (mean ± std.ev. across all subjects and voxels, max 0.98 ± 0.22 % across subjects) of variance. These results suggest that human RSC at most only weakly represents head direction during active navigation, and that it is not a consistent effect across subjects.

3.4. Discussion

In this experiment, we developed an immersive and interactive navigation experiment for fMRI, and fit an unprecedented number of models to brain activity collected from this experiment. We were able to simultaneously capture many navigational representations using a single dataset, and also test multiple hypotheses about representations. Results show that active navigation elicits activity across the cortex, and engages multiple distinct brain networks. These networks are largely consistent across subjects, and appear bilaterally symmetric.

This is not the first neuroimaging experiment to use an active navigation task in a naturalistic environment. (Spiers and Maguire, 2006) had also recorded brain activity while subjects drove in a virtual city. Indeed, that experiment, in one respect, is superior to this

experiment in that it used a real-world location (London). However, that study was limited by its contrast-based analysis methods and produced only maps of cortical activation during different cognitive states. It did not produce any quantitative models of how and what navigational features are represented in brain activity. In previous studies, we had shown voxelwise modeling to be effective for producing quantitative models of cortical representations from data from experiments using naturalistic stimuli (Deniz et al., 2019; Huth et al., 2012; Matusz et al., 2019). While these experiments had used more naturalistic stimuli than traditional fMRI experiments, they had used passive tasks and fixed stimuli. In this study, we showed that voxelwise modeling can also effectively analyze data from active tasks in which subjects directly interacted with the stimuli. We produced individualized models of how multiple types of navigational information are represented during active driving.

In addition to demonstrating that VM can enable the analysis of complex experiments, we also demonstrated its utility for testing hypotheses and replicating other results. Here, we tested 5 exemplar hypotheses about representation drawn from both the neuroimaging and neurophysiology literatures. In traditional experimental paradigms, each hypothesis would have necessitated a separate experiment. VM enables us to test them simultaneously, and in a more naturalistic setting than permitted by traditional experiments. A more naturalistic experiment better replicates real-world conditions, which is critical for inherently active tasks such as navigation. VM is also more informative than traditional point-null hypothesis testing frameworks. In addition to statistical significance, we are also able to determine the amount of variance each model explains in each voxel. Explained variance provides an estimate of effect size, and allows for the comparison of the relative contribution of multiple representations to brain activity. VM has enabled us to both replicate finding from more controlled neuroimaging studies, and also test the generalization of results from animal studies to the human brain. Crucially, VM has allowed us to close one of the critical gaps identified in section 1.4: the disconnect between the neuroimaging and the neurophysiology literatures.

The p-values for the hypotheses testing were computed with permutation tests. The null distribution was computed using ridge models fit using only the corresponding feature space. Stimulus correlations will inflate the performance of these ridge models. Thus, these p-values reflect a conservative estimate of model performances.

One limitation to these results is the lack of a functional noise ceiling. In experiments with fixed stimuli, it is possible to repeat the same stimulus multiple times. A noise ceiling can then be calculated by averaging these repeats, assuming that brain activity in response to the stimulus is stationary. However, in an active paradigm such as driving, it is impossible to present the same stimulus twice. Although the simulator can be configured to present the same situations, subjects cannot reliably perform the same action sequence identically. Because the virtual world is a complex dynamic system, small deviations will cause the conditions to quickly diverge across repeats. Furthermore, active navigation depends on many internal cognitive processes. Repeats of the same stimulus are not guaranteed to elicit the same navigational representations if the subject adopts a different plan.

It is possible to present subjects with multiple repeats of pre-recorded active driving videos in an effort to determine a proxy noise ceiling. However, we do not believe that to be a valid approach. Watching a pre-recorded video is inherently a different task than active navigation. Given results from both animal and human studies, there will be significant attentional and cognitive differences. The noise ceiling calculated from the passive video-watching task will not be applicable to active navigation. In a pilot control experiment, a subject watched multiple repeats of pre-recorded videos from active driving. A noise ceiling was computed form this data. We had found that using this noise ceiling to normalize model performances in active navigation resulted in many voxels being assigned an EV greater than 1. This undefined behavior led us to abandon this proxy noise ceiling. We take this lack of noise ceiling as a necessary trade-off in the next evolution of naturalistic experimental paradigms. Thus, all data have been presented here with a theoretically noise ceiling of 1.

Stimulus correlations are a significant concern in naturalistic experiments such as this one. As experiments becomes less controlled and more complex, stimulus correlations become stronger. Misattribution of variance to models may lead to erroneous conclusions about representations. Banded ridge regression provides a method for partitioning variance between multiple models. However, it is not a foolproof method for overcoming stimulus correlations. Correlated feature spaces must be all included during model fitting for variances to be correctly assigned to models. As we demonstrated in the case study, when feature spaces are omitted during model fitting, performances of models based on correlated feature spaces are artificially inflated. This misattribution of variance can significantly change conclusions. In the most extreme case, it can lead to erroneous conclusions. Many neuroimaging studies only examine a single set of features, and are thus susceptible to this issue. Therefore, it is critical to build feature spaces to capture as many aspects of the stimuli and task as possible.

In this experiment, we constructed 33 different feature spaces that span many different types of navigational information. However, we do not claim that they form an exhaustive list of all possible navigational features and their parametrizations. These feature spaces do not encompass all representations elicited by the active driving task, and there may also be better parameterizations of existing feature spaces. Nevertheless, these feature spaces have already accounted for many stimulus correlations, and have produced quantitative and predictive models of cortical representations of navigation in a complex interactive task. To our best knowledge, these models form the most comprehensive description to date of cortical navigational representations of navigational in the most naturalistic setting possible in fMRI. Thus, they are also the least biased models to date. These models provide a baseline and reference point from which new feature spaces and models could be developed to study human navigation in a naturalistic setting.

4 Active navigation elicits visual semantic tuning shifts across the cortex

4.1. Introduction

In the natural world, humans and animals dynamically switch between multiple tasks. These tasks are diverse in nature and changes the relative importance of different types of information. For example, during navigation, buildings that can serve as landmarks are more important, while in social situations, facial identities become more important. Thus, tasks may elicit different attentional states.

Attention is a powerful mechanism to increase the brain's efficiency for processing task-relevant information. Attention causes tuning shifts across the cortex, and at multiple levels of information processing. Non-human primate neurophysiology shows that in early visual areas, spatial attention causes shifts in the both the baseline and gain of neurons (Luck et al., 1997; McAdams and Maunsell, 1999; Reynolds et al., 2000), and can also shift their spatial receptive fields towards the attended location (Connor et al., 1997). Human neuroimaging show that across the cortex, visual semantic attention causes shifts in the semantic tuning of voxels towards the attended category (Çukur et al., 2013). These attentional warping of representations reflect a matched-filter mechanism that improves task performance by expanding the cortical representation of the attentional targets (David et al., 2008; Mazer and Gallant, 2003). This expansion in representation allocates more resources towards the targets, improving either resolution or processing speed.

Task-related tuning shifts have been demonstrated in electrophysiology in multiple rodent brain regions. In a landmark-based navigation task, rat RSC neurons are strongly tuned to landmarks that were indicative of reward location (Fischer et al., 2020). However, when rats saw the same landmarks outside of the task, the same neurons did not respond. In another experiment, rats performed a visually-guided task, a memory-guided task, and an

accumulating evidence task that all used the same stimuli (Pinto et al., 2019). The more difficult tasks elicited more decorrelated activity across rat dorsal cortex, suggesting that there were task-related tuning shifts across the cortex. In a third experiment, rats were directed to either freely forage or navigate to a memorized reward location (Ledergerber et al., 2021). In both tasks, neurons in the rat subiculum are selective for conjunction of multiple features, including position, head direction, and speed. However, the number of neurons with conjunctive selectivity was significantly higher in the goal-directed task. These studies show that there are changes in neural tuning across the cortex when animals switched tasks.

However, evidence of task-related tuning shifts is less clear in non-human primates or humans. In such experiments, attention is typically explicitly manipulated in a task. For example, in a spatial attention experiment, subjects may be cued to attend to different locations on the screen (Connor et al., 1997). Neurons shift their spatial tunings towards the attended location. In a visual semantic attention experiment, subjects may be cued to attend to particular object categories (Çukur et al., 2013; David et al., 2008). Voxels and neurons shift their semantic selectivities towards the attended category. However, these attentional shifts occur within the context of a single task. To our knowledge, there has been few studies on tuning shifts between tasks.

Here, we examine the difference in visual semantic tuning between a passive movie watching task and our active navigation task. In a previous experiment from this lab, subjects passively watched naturalistic videos. We then used voxelwise modeling to characterize the visual semantic tuning across the cortex in the two experiments (Çukur et al., 2013; Deniz et al., 2019; Huth et al., 2012, 2016; Lescroart and Gallant, 2019; Naselaris et al., 2011; Nishimoto et al., 2011). We then created maps of visual semantic tuning in each individual subject in passive movie watching and active navigation. Comparison of the visual semantic models and maps in two tasks provide a sensitive and objective method for determining how visual semantic selectivity changes due to task demands.

4.2. Methods

4.2.1. Passive movie watching experiment

The passive movie data was collected following the procedures outlined in (Nishimoto et al., 2011) and (Huth et al., 2012). Briefly, natural movies were drawn from the Apple QuickTime HD gallery and Youtube, and cropped to 512 x 512 pixels. Multiple 10-20 second clips were concatenated to form 21 10-minute runs. 12 runs were used as model training runs, and 9 runs were used for model validation. Four subjects (3 male, 1 female, ages 25-32) watched these movies at fixation while BOLD activity was recorded.

To create the semantic labels, an observer manually tagged every second of the movies with WordNet labels (Huth et al., 2012; Miller, 1995). Categories were tagged if they appeared in at least half of each second. Binary indicator features were then constructed for all categories that appeared in the movies. Additionally, features were also included for all superordinate categories for the labeled categories. For example, in addition to “dog,” features were also

Figure 4.1 The gaze semantics feature space captures the current visual semantic categories around the fixation point. During MRI data collection, subjects are allowed to freely gaze around the screen, and eyetracking is recorded at 60 Hz. To construct the gaze semantics features, we first use the game recordings to produce a ground-truth semantic segmentation of the scene on every frame into 16 navigationally-relevant categories. Eyetracking data is then used to determine the gaze position on each frame. A circle of 5° is then selected around the gaze position. Feature values are then calculated as the fraction of this circle each category occupies. Framewise features are then downsampled to match the sampling rate of the MRI scanner.

included for "canine" and "domestic animal." These total 1705 features. One additional feature was included to capture the total motion-energy power of the video at each second. Features were then downsampled to match the sampling rate of the scanner.

Voxelwise models were fit to data in each subject using ridge regression. To capture the hemodynamic response, an FIR filter was implemented by including delayed copies of the features. These copies were delayed by 2, 3, and 4 TRs (4, 6, and 8 seconds). The optimal weight parameters were selected by cross-validation over the training runs, and models were then evaluated by predicting brain activity on the validation set. For each voxel, a semantic tuning vector was then created by taking the average weight across delays for each feature.

4.2.2. Active navigation task

The active navigation data was collected as detailed in section 3.2. Briefly, subjects used a MR-compatible steering wheel and pedal set to drive in a virtual world built with Unreal Engine 4 and CARLA (Dosovitskiy et al., 2017). Prior to scanning, subjects learned the layout of the world. In the scanner, subjects performed an active taxi-driver task while BOLD activity was recorded. On each trial, subjects were cued to navigate to particular locations in the world. They then drove there via the shortest path, and must respect all traffic rules. After subjects arrive at the destination, a new trial begins. Subjects were allowed to freely view the screen, and eyetracking data was collected at 60 Hz. Trials were recorded with Unreal Engine's built-in demo recording system. Data was collected in 11-minute runs. We collected 11 runs from subject 1, and 18 runs from subject 2 and 3.

4.2.3. Visual semantic features in active navigation

Visual semantic features from the subject's perspective at each frame was then extracted from the game recordings (Figure 4.1). We first use the game recordings to produce a ground-truth semantic segmentation of the scene at 15 fps into 16 navigationally-relevant categories (See section 3.2.9.5). Eyetracking data is then used to determine the gaze position on each frame. A circle of 5° is then selected around the gaze position. Feature values were then calculated

as the fraction of this circle each category occupies. These values were first calculated at 60 fps, the eyetracking sampling rate. Framewise features were then downsampled to match the sampling rate of the MRI scanner.

4.2.4. Regression procedure

We then fit models to the BOLD activity using these visual semantics features. To account for the hemodynamic response, we fit a finite impulse response filter to each feature in each voxel. The FIR filter is implemented by including features are 5 temporal delays (0, 1, 2, 3, and 4 TRs). We then regressed the features $\boldsymbol{X}$ on to the BOLD activity $\boldsymbol{Y}$ using ridge regression. Ridge regression models brain activity in each voxel as a linear combination of the features, subject to a zero-mean spherical Gaussian prior on the weight magnitudes. The optimal solution to ridge regression is expressed as

$$\hat{W} = \underset{W}{\operatorname{argmin}} \tfrac{1}{2}\|Y - XW\|_2^2 + \|\lambda W\|_2^2$$

in which $\boldsymbol{\lambda}$ is a regularization hyperparameter. In our ridge solution, a single $\boldsymbol{\lambda}$ is optimized over the entire brain and applied to all voxels. This global $\boldsymbol{\lambda}$ ensures that weights are scaled equally across all voxels, and enables us to compare tuning vectors across voxels.

The optimal $\boldsymbol{\lambda}$ is selected by performing 10-fold cross-validation twice on the training set. Model performances are then evaluated by predicting brain activity on a hold-out validation dataset. Correlation between the predicted brain activity and actual brain activity was used to quantify model performance. Semantic tuning vectors were then created by taking the average weight across delays for each feature in each voxel.

4.2.5. Comparison of semantic tuning across tasks

Because semantic tunings for passive movie watching and active navigation were collected from different subjects and modeled with different feature spaces, they need to mapped into a common space before tuning shifts could be determined.

4.2.5.1. Mapping between subjects

To map semantic tunings across subjects, we used freesurfer's surf2surf function (Dale et al., 1999). Semantic tuning vectors in all subjects were first mapped to vertices on their cortical surfaces. In this experiment, we treat passive movie watching as the baseline semantic tuning. Huth et al. 2012 had shown that semantic tuning is largely consistent across subjects, with 6-8 group PCs explaining 30%-35% of model weight variances. We therefore took the average tunings across the 4 passive viewing subjects. To do so, we used surf2surf to calculate a mapping from each passive subject's surface to the fsaverage surface. Each subject's semantic tuning vectors were then mapped to the fsaverage surface. Vertex semantic tuning vectors from the four passive viewing subjects were then averaged on the fsaverage surface.

Then, for each active navigation subject, we used surf2surf to compute a mapping from the fsaverage surface to that subject's surface. We then mapped the average passive visual semantic tuning vectors to the cortical surface of each active navigation subject. Doing so allows us to compare semantic tunings on a vertex level in the native space of each active navigation subject.

4.2.5.2. Comparing across visual semantic feature spaces

Semantic tunings must also be compared in the same semantic space. While the WordNet labels for passive movies comprised 1705 hierarchical categories, the semantic segmentation of the active navigation data comprised 16 mutually-exclusive categories. We therefore selected the largest semantic subspace spanned by the intersection of these two sets. We used the smaller feature space, the active navigation task visual semantics, as the starting point. For each of the 16 categories, we first checked whether it is present in the WordNet labels in the passive movie-watching task. If it is present, we included that category as a dimension in the common space. 12 categories were thus selected to form a common semantic subspace in which to compare semantic tuning across tasks. For each of the 12 selected categories, we took its WordNet weight, and also the weights for all its WordNet hypernyms. We then averaged these hierarchical WordNet weights to a single weight vector, and used this vector as that category's passive weight.

Four categories from the active driving were excluded. "None" and "other" could not be mapped into the WordNet labels. The "none" label was used for any object that was not semantically tagged, while "other" were used for as a catch-all category that did not fit into any of the other semantic labels. "Road lines" is an extremely driving-specific category that was not labeled in the passive movie watching task. "Self" is a category that did not appear in the passive movies due to the third-person nature of the passive stimulus.

4.2.6. Computing the semantic tuning shift between tasks

After mapping the passive semantic weights to the active navigation subjects, and finding a common semantic subspace between the feature spaces, we are able to compute the semantic tuning shifts between the passive movie watching and active navigation tasks. At each vertex in each subject, we compared the two 12-dimensional semantic tuning vectors, one from the passive movie watching task, and the other from the active navigation task. First, to determine the magnitude of the semantic shift, we computed the correlation between the two tuning vectors at each vertex. A correlation of 1 would indicate that there is no semantic tuning shift, while a correlation of -1 would indicate that the vertex shifts to opposite selectivities between the two tasks. A correlation of 0 would indicate that the vertex shifts to orthogonal selectivities between these two tasks.

Next, at each vertex, we also determined the direction of the semantic shift. At each vertex, we computed two unit vectors, one in the direction of the passive movie-watching semantic tuning vector, and the other in the direction of the active navigation semantic tuning vector. We then subtracted the passive movie-watching unit vector from the active navigation unit vector. This difference describes the tuning shift direction in the 12-dimensional semantic subspace at each vertex.

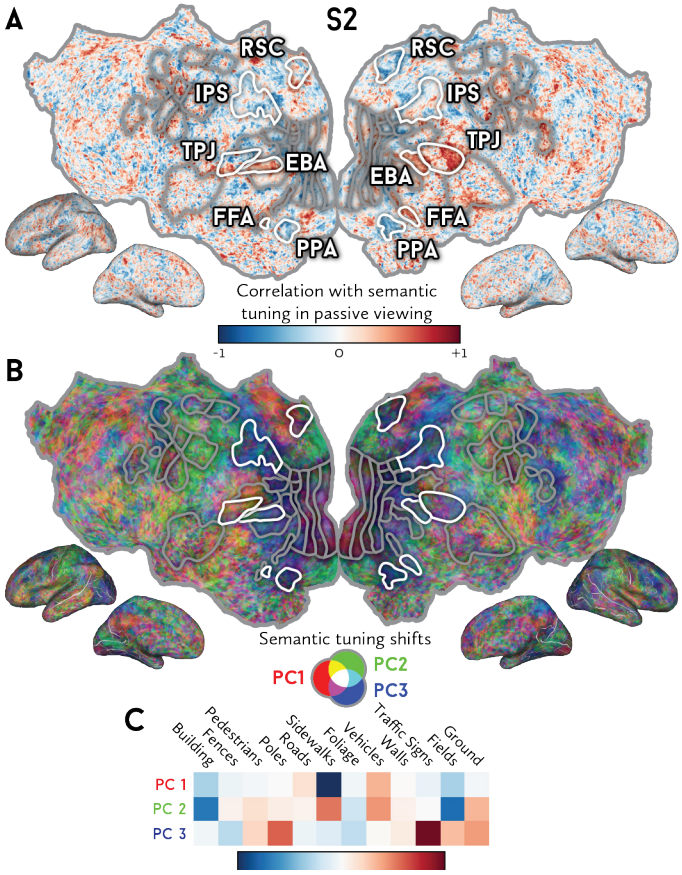


Figure 4.2 Active navigation induces visual semantic tuning shifts across the cortex. In each subject, we compared cortical tunings for 12 visual semantic categories in both this experiment and a passive movie watching experiment (Huth et al 2012). We used the group-level cortical semantic tuning vectors from the passive movie watching task. Here we show tuning shifts on the cortex for subject 2. A) We correlated semantic tuning between the active driving and passive movie tasks across the cortex in each subject. A correlation of 1 indicates no tuning shift, 0 indicates a tuning shift to an orthogonal selectivity, and -1 indicates a shift to the opposite selectivity. We find that there are significant tuning shifts across the cortex, particularly in RSC, OPA, PPA, and IPS. B) We calculated the tuning shift direction across the cortex, and used PCA to discover the most important dimensions in these shifts. We show the first three PCs on the red, green, and blue channels respectively. The direction of the tuning shifts also varies by region. C) We then examined these tuning shift PCs. The first PC shows a tuning shift away from sidewalks and towards vehicles. The second PC shows a tuning shift towards sidewalks, vehicles, pedestrians and the ground. The third PC shows a tuning shift towards traffic signs and poles, and also ground, fields, and pedestrians. These results that active navigation causes significant tuning shifts in semantic representation compared to passive movie watching.

4.3. Results

Rodent electrophysiology has shown that changing tasks cause tuning shifts across the cortex. This is likely due to inherent attentional demand differences between tasks. Attention causes significant tuning shifts across the cortex towards the attentional target. In most human studies, attentional effects are generally examined within a single task and not compared across tasks. To examine tuning shifts between tasks, we compared visual semantic

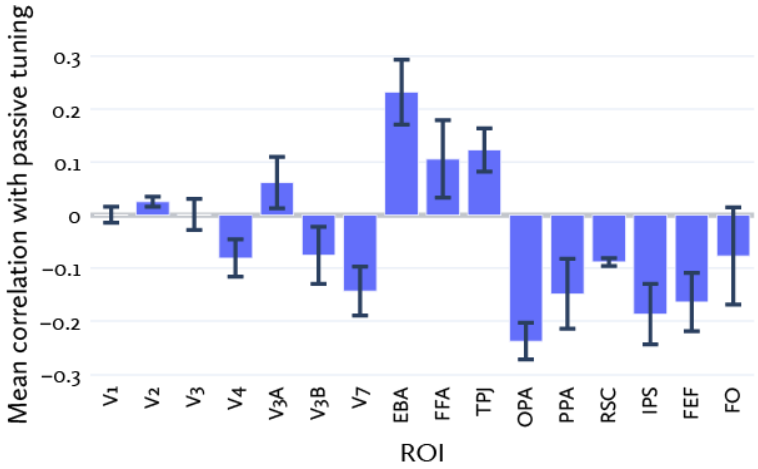


Figure 4.3 We examined the average semantic tuning correlation between passive viewing and active navigation in 16 ROIs. Error bars indicate standard errors of the mean across subjects. EBA, FFA, and TPJ were most correlated between the two tasks. In other words, they showed the least tuning shift between the two tasks. On the other hand, semantic tuning in OPA, PPA, RSC, IPS, FEF, and FO were negatively correlated between the two tasks. In other words, they showed much stronger tuning shifts, and shifted their tunings towards opposite categories in the two tasks. We also included more posterior retinotopic areas (V1, V2, V3, V3A/B, V4, and V7). These areas show an intermediate amount of semantic tuning shifts. However, these retinotopic regions are not as semantically-selective as anterior visual regions, and may suffer more from noise in this visual semantics model. These results show that visual semantic tuning shifts are not consistent across functional ROIs, and that visual navigation and attention areas display much stronger tuning shifts than other ROIs.

tunings between a passive movie-watching task and an active navigation task. In both tasks, we used voxelwise modeling to first capture the visual semantic tunings across the brain. When then quantified the difference in visual semantic tuning between the two tasks.

In all subjects, we found significant tuning shifts across the cortex between the two tasks. We show the tunings in one subject in Figure 4.2 (For other subjects see Supplementary Figures 7 and 8). The tuning shifts are not uniform across the cortex (Figure 4.2A). To better understand these tuning shifts, we used PCA to determine the most important tuning shifts. We concatenated all tuning shift directions across subjects, and applied PCA to the aggregate tuning shift direction vectors. In Figure 4.2B we plot the top 3 PCs of these tuning shifts, and show the weights of the 12 categories on these PCs. The first PC shows a strong shift away from the representation of sidewalks, and towards the representation of vehicles. The second PC shows a shift towards the representation of sidewalks, vehicles, and the ground, and weakly towards pedestrians. The third PC shows a strong shift towards the representation of traffic signs and poles (upon which traffic signs are mounted), and weakly towards pedestrians, fields, and the ground. Interestingly, the third PC of semantic tuning shifts dominates RSC, OPA, IPS, and PPA, the ROIs that show the strongest tuning shifts across tasks.

The magnitude of the tuning shifts varies across the cortex. Therefore, we also examined tuning shift magnitudes across ROIs (Figure 4.3). As this model is fit with a visual semantic feature space, we examined visually-relevant functional ROIs. Semantic tuning in EBA, FFA, and TPJ were most correlated between the two tasks (EBA: 0.23 ± 0.06 mean ± s.e.m. across

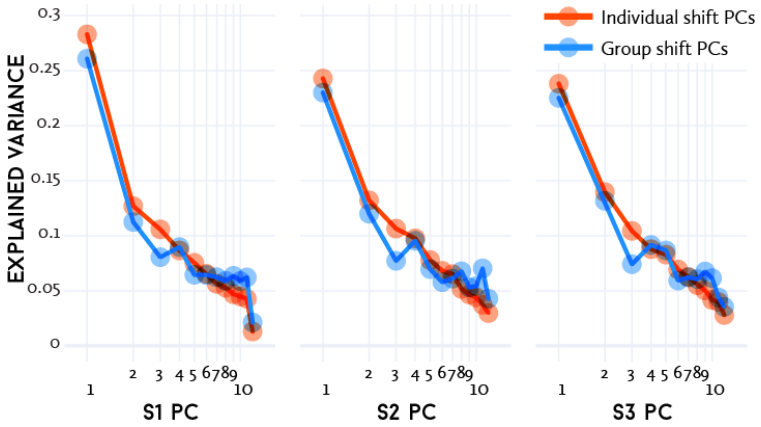


Figure 4.4 We used PCA to discover the primary components of semantic tuning shifts. Here, we show the amount of variance in semantic tuning explained by these PCs. For each subject, we took the cortical tuning shift directions of the other two subjects, and computed a group semantic shift direction PC space. Red lines show the amount of variance explained by each subject's own PCs. Blue lines show the amount of variance explained by group PCs calculated from the other two subjects. We find that the group PCs are able to explain roughly the same amount of variance as each subject's own PCs. This suggests that task-related shifts in semantic representation are consistently replicated across subjects.

subjects; FFA: 0.11 ± 0.07, TPJ: 0.12 ± 0.04). In other words, they showed the least tuning shift between the two tasks. On the other hand, semantic tuning in OPA, PPA, RSC, IPS, FEF, and FO were negatively correlated between the two tasks (OPA: -0.24 ± 0.03; PPA: -0.15 ± 0.07; RSC: -0.09 ± 0.01; IPS: -0.19 ± 0.06; FEF: -0.16 ± 0.05; FO: -0.08 ±0.09). In other words, they showed much stronger tuning shifts, and shifted their tunings towards opposite categories in the two tasks. We also included earlier, retinotopic visual areas in this analysis. These areas displayed shifts to mostly orthogonal tunings between the tasks, and tunings were largely uncorrelated (V1: 0.00 ± 0.02, V2: 0.03 ± 0.01, V3: 0.00 ± 0.03, V4: -0.08 ±0.04). However, earlier visual areas are less semantically tuned, and thus semantic tuning vectors may be much more susceptible to noise. Random samples of high-dimension vectors tend to be orthogonal to each other, and therefore these noisy tuning vectors are likely to be uncorrelated across tasks.

We also examined how well the tuning shift directions generalize across subjects. For each subject, we took the cortical tuning shift directions of the other two subjects, and computed a group semantic shift direction PC space. We then projected that subjects' cortical semantic tunings into this group PC space, and examined the amount of variance that the group PCs explained. We compared these to the amount of variance that subject's own PCs are able to explain (Figure 4.4). We find that the group PCs are able to explain roughly the same amount of variance as each subject's own PCs. This suggests that these tasks cause semantic tunings to shift in consistent directions across all subjects.

As a complementary analysis to the tuning shifts across the cortex, we also examined the cortical weights for each category in each subject. For every category, we computed the spatial correlation across the cortex between its passive and active representations in each subject (Figure 4.5). For many of the categories, their cortical weights were not correlated

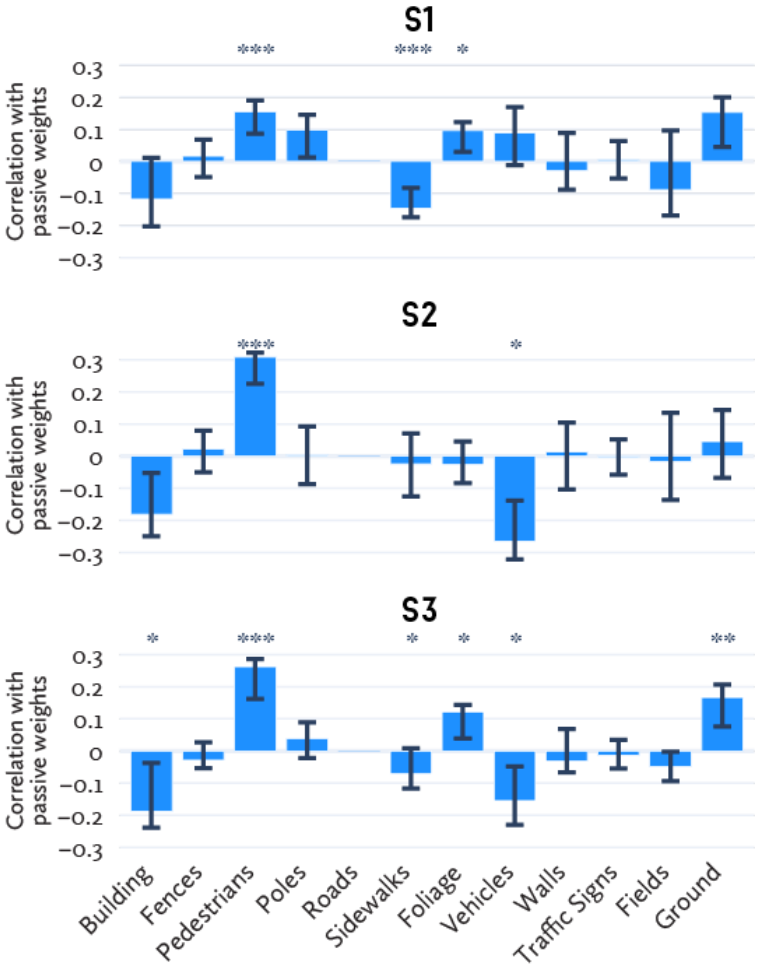


Figure 4.5 We examined the cortical representation to the 12 visual semantic categories, and computed the correlation of their cortical weights between passive movie watching and active navigation. These correlations reflect whether each category is represented by the same brain regions in the two tasks. Error bars indicate 95% confidence intervals. Stars indicate significance levels for correlations: ***: $p < 1e\text{-}4$, **: $p < 1e\text{-}3$, *: $p < 1e\text{-}2$, permutation test. Only one category, pedestrians, show a consistent representation across the two tasks in all subjects. In two out of three subjects, vehicles representations in active driving are represented in opposite regions than in passive movie watching. These results show that, for the majority of semantic categories we are able to examine in active driving, cortical tunings are shifted away from their representations in passive movie watching. Furthermore, the strength of these tuning shifts differs across categories.

across the two tasks ($p > 1e\text{-}2$, permutation test), suggesting that their representations shift to orthogonal networks of brain regions across tasks. Only one category, "pedestrians," show a correlated representation across the two tasks in all subjects ($p < 1e\text{-}4$, permutation test). In two subjects, "vehicles" was negatively correlated between passive movie watching and active driving ($p < 1e\text{-}2$, permutation test), suggesting that its representation shifted to opposite brain regions between the two tasks. Taken together, these data show that visual semantic tuning across the cortex significantly shifts across different tasks.

4.4. Discussion

In this experiment, we compared visual semantic tuning between a passive movie watching task and an active navigation task. We found significant tuning shifts across the cortex between the two tasks, and cortical representations are shifted in the same directions in all subjects. The magnitude of the semantic shifts varies across ROIs, with navigational and spatial attention areas showing the strongest changes. These results show that changing tasks induce significant shifts in visual semantic representation.

Post-hoc inspection of these tuning shifts suggest that they are reasonable in the context of these tasks. In both the 1st and 2nd PCs, there is a shift towards the representation of vehicles. This shift likely reflects the increased importance of vehicles in the active navigation stimulus versus that in passive viewing. This shift happens in the precuneus, TPJ, and prefrontal cortex, suggesting that vehicles engage more theory-of-mind and higher-level planning mechanisms during active navigation. In the 3rd PC of semantic tuning shift directions, we find a significant shift towards traffic signs and poles, upon which signs are mounted. In passive movie viewing, traffic signs are likely treated as background props. In active navigation, traffic signs are a critical component of the environment that govern the actions that subjects can take. This shift happens in visual navigation areas (RSC, OPA, and PPA), and also spatial attention areas (IPS and FEF), indicating that navigation areas are more selective for traffic signs, and subjects may also be attending more to these signs.

We recognize two limitations to the current study. First, the visual semantic feature spaces used in the two experiments do not span the same semantic space. In the passive movie task, object categories were labeled with WordNet categories by hand, and spanned 1705 categories. In the active navigation task, object categories are generated by the game engine, and span only 16 categories that are useful for navigation via driving. The active navigation semantic labels differ from the passive movie viewing labels in terms of both semantic resolution and also distribution. Resolution in some parts of semantic space is very coarse in active driving. For example, many items, such as trash cans, flower pots, and ornamental statues, are treated as the same "props" category. On the other hand, semantic resolution in navigation-specific categories is much finer. For example, the labeling differentiates road lines from the road surface, and traffic signs from the poles upon which they are mounted.

The frequency of different semantic categories also differs between the two tasks. For example, roads, sidewalks, and buildings are overrepresented in the active navigation stimulus, while animals are underrepresented. The active navigation labels also included the category of "self," which was not present in the passive movie watching stimuli. To ameliorate these discrepancies, we limited our analysis to only the semantic subspace spanned by both feature spaces. Thus, we may not have been able to fully capture all semantic tuning shifts between the two tasks. Nevertheless, we are able to demonstrate strong semantic tuning shifts even in this limited subspace.

Second, the subjects in the two tasks were different. The weights from the passive viewing subjects were mapped on to the fsaverage surface, averaged together, and then mapped to the active navigation subjects' cortical surfaces for comparison. The two mapping steps through

the fsaverage surface necessarily induce some resolution loss, as anatomical idiosyncrasies cannot be losslessly mapped between subjects. The averaging of weights from the passive subjects within fsaverage space also induces resolution loss. However, previous studies on semantic representation had shown that the first 5-6 semantic PCs are consistent and generalize across subjects. Thus, we believe that it is valid to use the group average passive semantic weights for computing semantic tuning shifts in at least that many dimensions.

These results form an extension of previous work on semantic tuning shifts. In previous experiments, we had shown that an attention-switching task, given the same stimulus, warps semantic representations across the cortex. Here we show that semantic tuning across the cortex also shifts significantly across tasks. These results suggest that semantic tunings are not stationary across tasks. This dynamic representation of visual semantic can improve the effective efficiency of the brain, by shifting limited neural resources towards task-relevant categories as the subject switches tasks. Overall, these findings help explain how the human brain is able to perform many diverse tasks in complex, ever-changing environments.

5 Representation of abstract route progression in anterior visual areas

5.1. Introduction

In goal-directed navigation, subjects travel towards a destination following a pre-planned route. As subjects navigate, they must keep track of their progression along the planned route. In rodents, neurons in the retrosplenial cortex (RSC) and parietal cortex have been demonstrated to represent route progression information. As rats ran along a track, parietal cortex neurons displayed spatial firing fields for distinct segments of the track (Nitz, 2006). These firing fields moved and changed size to align to conceptually identical portions of different paths. For example, one parietal cortex neuron fires selectively for the segment of routes before the first turn. When rats ran on a closed loop, some RSC neurons displayed spatial firing fields that were localized to particular locations on the track (Alexander and Nitz, 2017). For example, one RSC neuron may fire at a quarter of the way around the track. Other RSC neurons displayed spatially periodic firing fields. For example, another RSC neuron displays peaks spaced equally apart along the track. Like parietal cortex neurons, these RSC neurons retained their firing fields at conceptually identical positions on a differently-shaped track. In other words, these neurons represent position in an abstract reference frame defined by the route, rather than a position in a concrete reference frame such as the room. These results suggest that RSC and PPC neurons encode abstract information about position along a route, rather than pure spatial position.

Most studies of navigation implicitly assume that space is represented metrically (see section 1.4.1). In metric representations, locations are represented on a map that preserves their exact positions in the world. For example, place cell firing fields are anchored to specific locations in the world (O'Keefe and Dostrovsky, 1971; O'Keefe and Nadal, 1978), and grid cells fire on a spatially regular lattice (Brun et al., 2008; Fyhn et al., 2004; Hafting et al., 2005). However, it is also possible to represent space topologically. In topologic representations, locations are

represented on a graph that preserves only their relationships (Chrastil and Warren, 2014). These rodent neural representations of position along a route are warped to align with the physical shape of the track. In other words, route progression is represented topologically rather than metrically, as the firing fields are aligned only to relative positions, rather than absolute positions.

Rodents are used as a model organism for human navigation. Human electrophysiology studies in epilepsy patients have demonstrated that the human hippocampus and entorhinal cortex share the same metric spatial representations with rodents (Ekstrom et al., 2003; Jacobs et al., 2013). Namely, these are place cells and grid cells. However, homologies between humans and rodents are less clear in the neocortex than in subcortical structures. Thus, neocortical representations may also diverge between the species, and it is critical to test whether results from rodents generalize to humans.

Functional MRI non-invasively records brain activity (Ogawa et al., 1990), and has enabled experiments in healthy human subjects. However, to our knowledge, many navigational representations found in rodents, including route progression, have not been examined in humans. Results from rodents are often acquired from actively moving animals. Because of the nonlinear nature of the brain, it is likely that many navigational representations are only present during active navigation (Fischer et al., 2020). Indeed, rodent experiments often exclude data during which the animals are stationary. However, most previous human navigation experiments used passive stimuli and did not have an active navigation task. Furthermore, most common analysis methods are ill-suited for analyzing dynamic brain activity in active tasks. Thus, most human neuroimaging navigation experiments could not test these rodent representations seen during active navigation.

In this experiment, we overcome the limitations of previous human neuroimaging experiments. We recorded BOLD activity while subjects performed an active navigation task in a naturalistic virtual environment. We then used voxelwise modeling (Çukur et al., 2013; Deniz et al., 2019; Huth et al., 2012; Lescroart and Gallant, 2019; Naselaris et al., 2011; Nishimoto et al., 2011) to map the representation of route progression across the cortex. Voxelwise model weights were then used to examine tuning to routes in anterior visual navigation regions.

5.2. Methods

5.2.1. Subjects

Three healthy adult volunteers (1 female) with normal or corrected to normal vision participated in this study: S1 (age 26), S2 (age 27), and S3 (age 32). The experimental procedures were approved by the Institutional Review Board at the University of California, Berkeley, and written informed consent was obtained from all subjects.

5.2.2. Active navigation task

The active navigation data was collected as detailed in section 3.2. Briefly, subjects used a MR-compatible steering wheel and pedal set to drive in a virtual world built with Unreal Engine 4 and CARLA. Prior to scanning, subjects learned the layout of the world. In the scanner, subjects performed an active taxi-driver task while BOLD activity was recorded. On each trial, subjects were cued to navigate to particular locations in the world. They then drove there via the shortest path, and must respect all traffic rules. After subjects arrive at the destination, a new trial begins. Subjects were allowed to freely view the screen, and eyetracking data was collected at 60 Hz. Trials were recorded with Unreal Engine's built-in demo recording system.

5.2.3. Scanning procedure

MRI data were acquired on a 3T Siemens Trio at the University of California, Berkeley with a 32-channel head coil. BOLD data were acquired with a custom T2*-weighted gradient-echo EPI sequence. This sequence uses a water-excitation radiofrequency pulse to prevent contamination from fat signal. The following parameters were used: TR = 2 s, echo time = 34 ms, flip angle = 74°, voxel size = 2.24 × 2.24 × 3.5 mm3, field of view = 224 × 224 mm2, matrix size = 100 × 100, and 32 axial slices. Custom personalized headcases (Power et al., 2019) were used to stabilize the head and reduce motion artifacts. To reconstruct cortical surfaces, we collected anatomical data with a three-dimensional T1-weighted MP-RAGE sequence that had the following parameters: 1 × 1 × 1 mm3 voxel size, 256 × 212 × 256 mm3 field of view.

BOLD was collected in 11-minute runs. We collected 11 runs of data from subject 1, and 18 runs of data each from subjects 2 and 3.

Eyetracking data was collected using an Avotec dark-pupil IR eyetracker at 60 Hz, and processed with custom software to extract gaze locations. At the beginning of every functional run, 35 calibration points were presented for 2 seconds each to subjects to ensure accurate eyetracking calibration. The taxi driver task began after the calibration sequence.

5.2.4. Route progression features

We created a set of route progression features that capture the subject's position along any route (Figure 5.1). On each trial, we first linearized the route that the subject took. To account for routes of different lengths, we then normalized all routes to a length of 1. We then divide this normalized route into multiple sets of bins. These bins divided up the route into 1, 2, 4, 8, 12, and 16 segments. On each frame, we then determined the subject's progression along this route. For each set of bins, we then determined the bin that the subject was in, and set that bin's feature value to 1. All other bin had values of 0. Thus, at any point on a route, there were 6 features with a value of 1, and all others had a value of 0. Note that the 1-bin set serves as simply an indicator that the subject is active navigating (and not in the intertrial delay period).

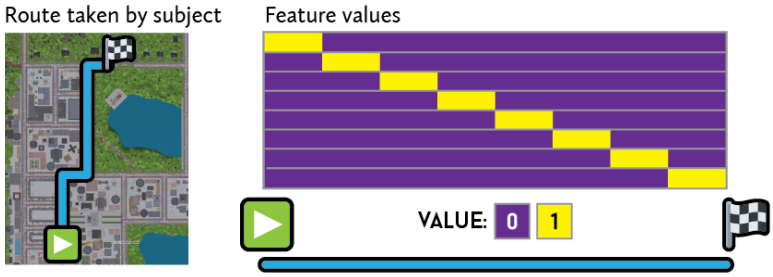


Figure 5.1 The space binned route progression feature space uses indicator variables to capture an abstract representation of the position of the subject along a planned route. On each trial, the subject's route (left) is first linearized and then normalized to a length of 1 (right). The normalized route is then divided into 4, 8, 12, and 16 bins, each of which corresponds to a single feature. Here we only show 8 bins for clarity. When the subject is in the corresponding portion of the route, that feature's value is 1, otherwise, its value is 0. During the intertrial delay, all features are zero. The same set of bins are used to capture position along the route in all trials, and thus reflect an abstract representation of position along a route that is not anchored to any particular location.

5.2.5. Regression procedure

We then fit models to the BOLD activity using these route progression features. To account for the hemodynamic response, we fit a finite impulse response filter for each feature in each voxel. The FIR filter is implemented by including features at 5 temporal delays (0, 1, 2, 3, and 4 TRs). Ridge regression was used to map features $\boldsymbol{X}$ on to BOLD activity $\boldsymbol{Y}$. Ridge regression models brain activity in each voxel as a linear combination of the features, subject to a zero-mean spherical Gaussian prior on the weight magnitudes. The optimal solution to ridge regression is expressed as

$$\hat{W} = \underset{W}{\operatorname{argmin}} \tfrac{1}{2}\left\|Y - XW\right\|_2^2 + \left\|\lambda W\right\|_2^2$$

in which $\boldsymbol{\lambda}$ is a regularization hyperparameter. In our ridge solution, a single $\boldsymbol{\lambda}$ is optimized over the entire brain and applied to all voxels. This global $\boldsymbol{\lambda}$ ensures that weights are scaled equally across all voxels. Equal scaling enables us to compare tuning vectors across voxels.

The optimal $\boldsymbol{\lambda}$ is selected by performing 10-fold cross-validation twice on the training set. Model performances are then evaluated by predicting brain activity on a hold-out validation dataset. Correlation between the predicted brain activity and actual brain activity was used to quantify model performance.

5.2.6. Route progression tunings

We then analyzed the weights for these spatial bins. For each feature, we averaged its weights across delays to obtain its mean tuning in each voxel. From these tunings, we constructed a route selectivity profile in each voxel. First, we divided the normalized route into 48 bins (the lowest common multiple of 1, 2, 4, 8, 12, and 16). Then, for each set of bins used to create the features, we mapped the bins to their corresponding sizes in these 48 bins. For example, bin 1 in the 4-bins division is mapped to bins 1-12 in these 48 bins. Weights are then summed for the 4-, 8-, 12-, and 16- bin divisions. Features from the 1- and 2-bin divisions were excluded because they effectively provided a constant value. This profile describes the tuning of each voxel to normalized position along any route.

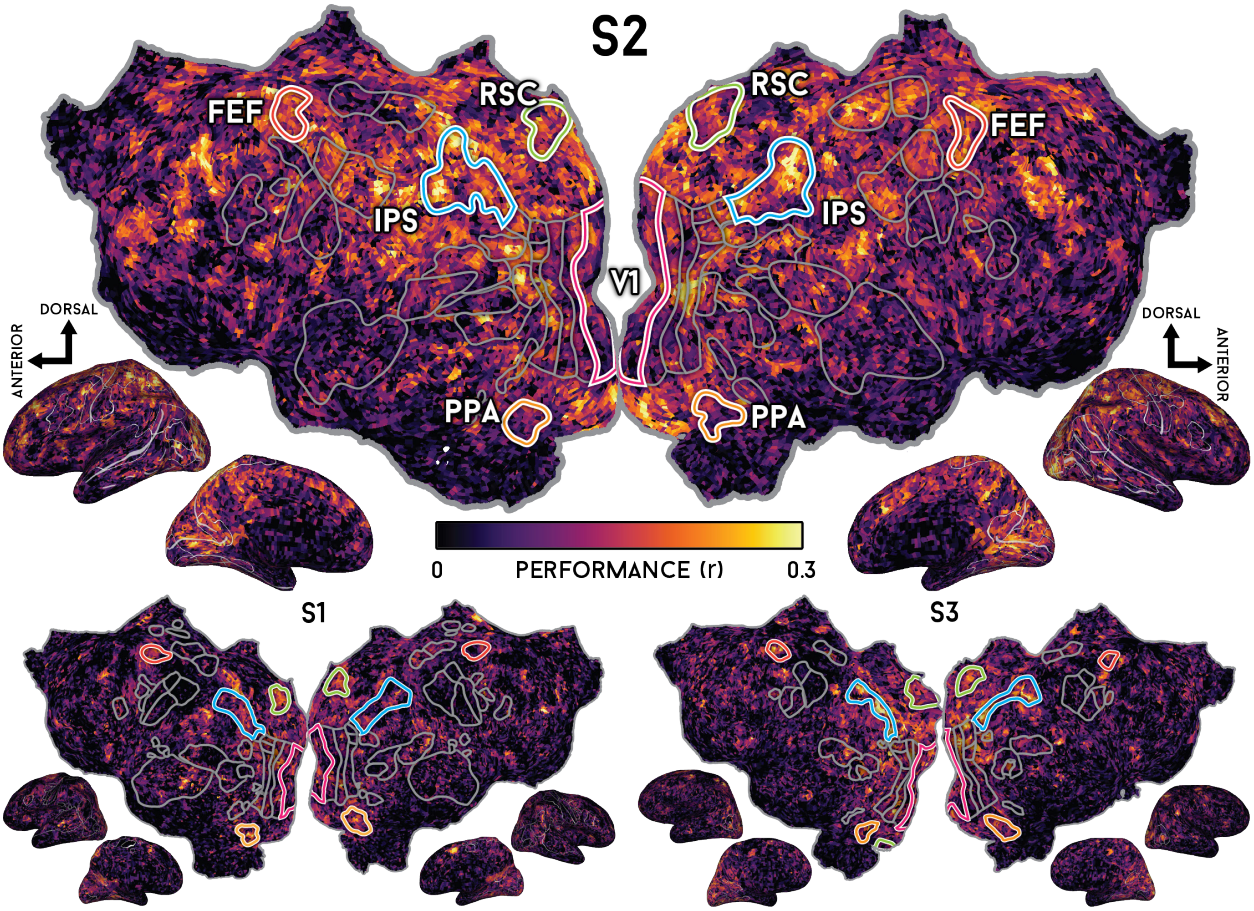


Figure 5.2 Model performance of the space binned route progression model. The binned route progression model allows us to examine the cortical tuning to portions of routes. We used ridge regression to fit a standalone model to the space binned route progression features. Here we show the model performance of the space binned route progression model across the cortex is the same layout as figure 3.1. Space binned route progression predicts well in anterior visual areas, including RSC, OPA, PPA, and IPS, and also in the precuneus, and parts of the prefrontal cortex. This result suggests that, similar to the rodent brain, the human brain also indeed represent abstract position along a route.

We also define a start-end preference value as a summary scalar description of these tuning profiles. In each voxel we first summed the weights for the first bin in each of the 4-, 8-, 12-, and 16-bin divisions. Again, the 1-, and 2-bin division features were excluded because were effectively constants. Next, we summed the weights for the last bin in each of the divisions. The start-end preference of each voxel is then defined as the difference between the sum of last bins and the sum of first bins. Thus, voxels with a negative start-end preference are more strongly tuned for the start of routes, and voxels with a positive star-end preference are more strongly tuned for the end of routes.

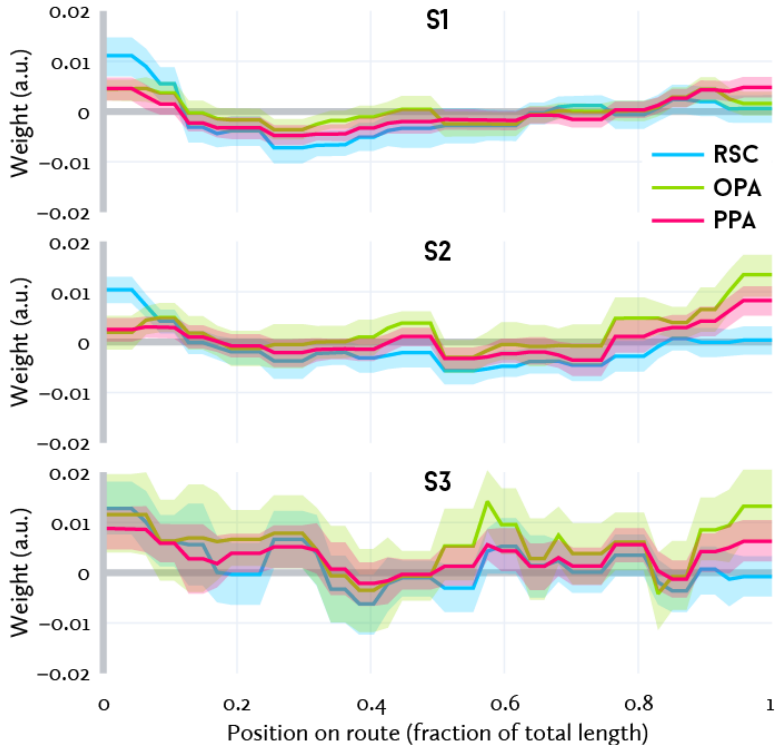


Figure 5.3 Mean selectivity to position on route in RSC, OPA, and PPA. Here we show the average tuning in RSC (blue), OPA (green), and PPA (red) for position along a route. Shaded areas indicate 95% confidence intervals. We find the strongest weights are the beginning and end of routes, while the middle of routes tends to have no weight or negative weight. Across all three subjects, we consistently find that while RSC is selective for the beginning of routes, it is not selective for the end.

5.3. Results

Electrophysiology studies in actively navigating rodents demonstrated that neurons in the RSC and PPC represent position along a route (Alexander and Nitz, 2017; Nitz, 2006). Neurons preserve their tuning to route progression across different routes, suggesting that it is an abstract representation that is not purely spatial. In this experiment, we examined whether the human brain also contains such representations. We recorded BOLD activity from human subjects actively navigating in a virtual environment. Voxelwise modeling was then used capture any route progression representations, and we examined the tunings of individual voxels to position along a route.

We find that the spatial binned route progression model predicts well in anterior visual areas, including RSC, OPA, PPA, and IPS (Figure 5.2). Additionally, this models also predicts well in anterior precuneus and parts of the prefrontal cortex. This pattern is consistent across subjects, suggesting that humans indeed also have the same representation of route progression as rodents. Furthermore, because route progression can be found in regions homologous to rodent RSC and PPC, these representations may be conserved across species. These representations capture route progression in an abstract manner, showing that these large-scale routes are represented topologically.

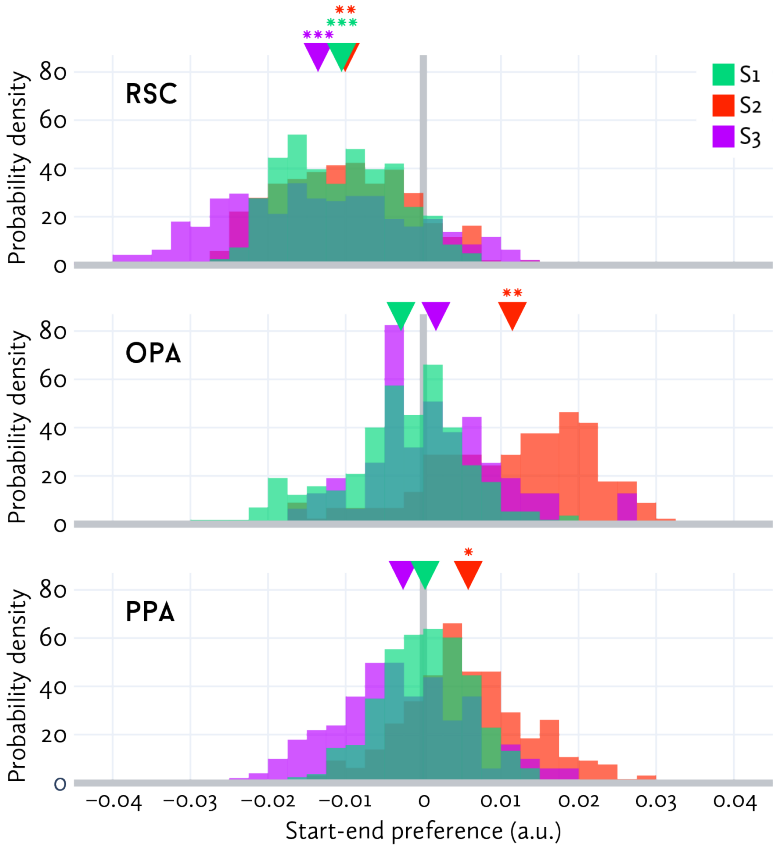


Figure 5.4 RSC consistently prefer the start of routes. For each voxel, we quantified its start-end preference by subtracting its weight for the first bin (route start) from that of the last bin (route end). Thus, negative values indicate a preference for the beginning of routes, 0 indicates no preference, and positive values indicate a preference for the end of routes. Here, we show histograms of start-end preferences in each ROI in each subject. Red corresponds to voxels in subject 1, blue to voxels in subject 2, and green to voxels in subject 3. Colored triangles indicate the mean start-end preference in each ROI for that subject. Stars indicate p-values for whether the mean start-end selectivity is different from 0: ***: $p < 1e\text{-}4$, **: $p < 1e\text{-}3$, *: $p < 1e\text{-}2$, permutation test. We find that RSC consistently prefers the start of routes in all subjects. However, there is no consistent preference in OPA or PPA across subjects. Both OPA and PPA preferred the end of route in a single subject (S2), but no preference in the other two subjects. These results suggest that RSC has a route progression tuning that is distinct from the other two ROIs.

Next, we examined the route progression tuning profiles in RSC, OPA, and PPA. We focused on these anterior visual areas because of their importance in human navigation. In Figure 5.3, we plot the average tuning profiles for each ROI in each subject. These ROIs tend to display the strongest tunings to the start and end of routes. On the other hand, the middle of routes tends to nave no weight or negative weight, suggesting that these ROIs are predominantly selective for starts and ends. However, these are the mean route selectivity profiles, and individual voxels may display different tunings.

We note that across all subjects, RSC is strongly selective for the start of routes, but not the end. Therefore, we also examined the start-end preferences of all voxels in RSC, OPA, and PPA. In Figure 5.4, we plot the distribution of start-end preferences in each voxel in each

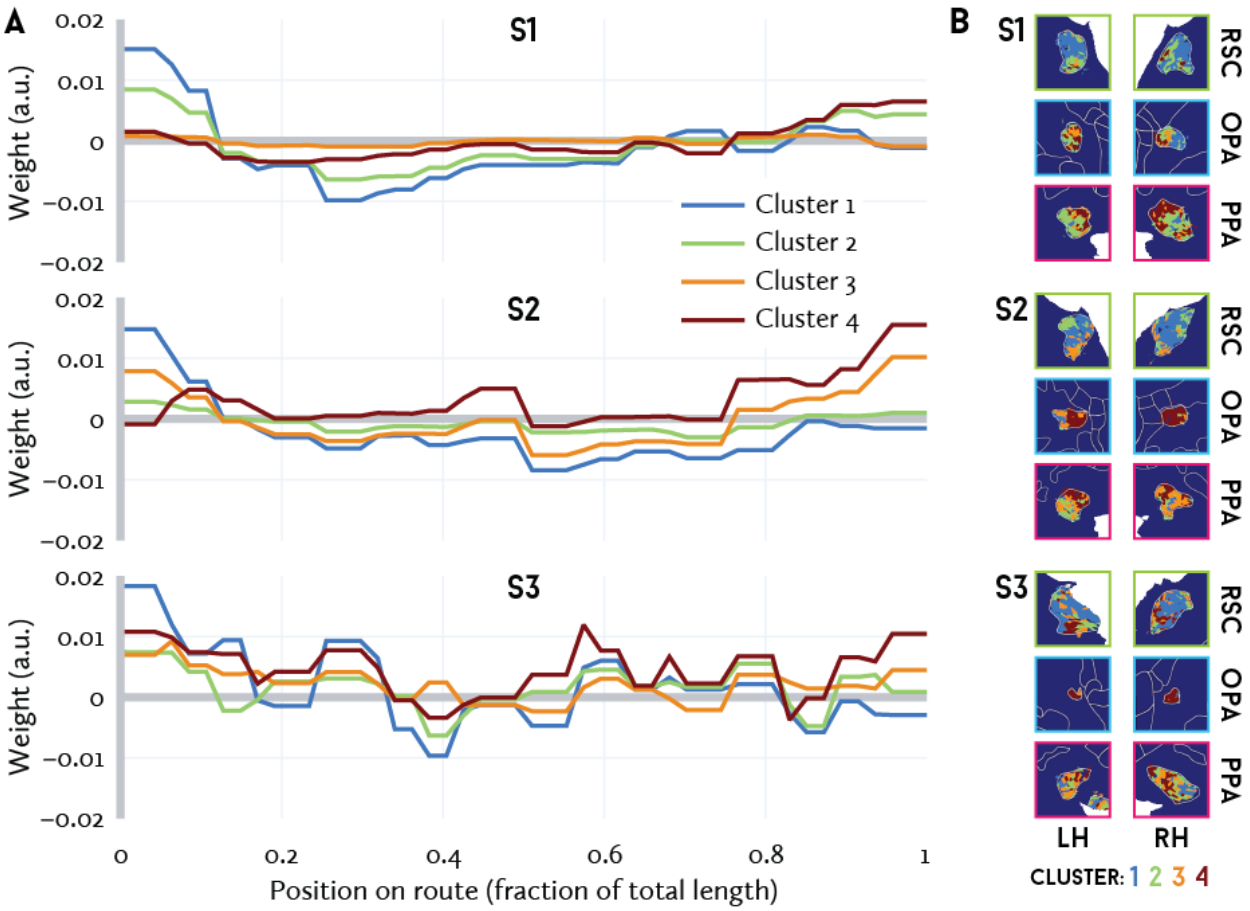


Figure 5.5 Unsupervised clustering isolates RSC. The start-end preference used only the first and last bins, and discarded weights from all other bins. Thus, RSC's distinction as a consistent start-preferring ROI may be an artifact from the selection of only these bins. To examine this distinction in an unbiased manner, we applied k-means clustering to the route position tuning profile of all voxels. In each subject, we clustered the voxels into four clusters. A) We plot the average route progression tuning profile for each cluster in each subject. There is at least one cluster in each subject that displays a strong selectivity for the beginning of a route, and little selectivity for the end of the route. Clusters are sorted by their average start-end preferences: cluster 1 is most start-preferring, and cluster 4 is most end-preferring. B) We then visualized these clusters on the cortex. Clusters are numbered and colored in the same order as in A. We find that in each subject, RSC is overwhelmingly assigned to cluster 1, the most start-preferring cluster. This unbiased, data-driven clustering of voxels show that RSC indeed has a different tuning profile to route progression than OPA or PPA.

ROI. Across all three subjects, RSC voxels consistently prefer the start of routes in all subjects ($p < 1e\text{-}4$, permutation test). However, there is no such consistent preference in the other two ROIs. In subject 2, OPA ($p < 1e\text{-}3$, permutation test) and PPA ($p < 1e\text{-}2$, permutation test) both prefer the end of routes. In the other two subjects, on the other hand, OPA and PPA do not show any significant preferences. These results suggest that RSC has a distinct route progression tuning profile that prefers the start of routes.

Because the start-end preference metric uses only weights from the first and last bins, it discarded much information about the tuning profile. Thus, RSC's distinction based on this metric may be an artifact from the definition of this metric. To examine this distinction

in a more unbiased manner, we applied unsupervised clustering to the route progression tuning profiles of all voxels. We used k-means clustering to divide all RSC, OPA, and PPA voxels into 4 clusters in each subject. In every subject, we find a cluster that is strongly start-preferring (Figure 5.5A). We then visualized the distribution of these clusters on the cortical surface (Figure 5.5B). We sorted the clusters by their average start-end preferences, with the most start-preferring cluster as the first cluster. In all three subjects, we find that cluster 1 is predominantly localized in RSC. Additionally, RSC voxels in all three subjects were mostly assigned to cluster 1. On the other hand, voxels in PPA and OPA were more heterogenous, though OPA in two subjects were more likely to be assigned to cluster 4, the most end-preferring cluster. This unbiased, data-driven clustering of voxels show that RSC indeed has a distinct tuning profile than OPA or PPA.

5.4. Discussion

In this experiment we bridge one of the key issues identified in section 1.4: the disconnect between neurophysiology results in rodents and neuroimaging results in humans. Rodent neurophysiology studies have shown that RSC and PPC contains an abstract representation of progression along routes. Here, we demonstrate that the human brain contains similar representations. We recorded BOLD activity while subjects actively navigated in a virtual environment. Voxelwise modeling was then used to map representations of route progression. Well-predicted regions include areas homologous with rodent regions that represent these features. This result demonstrates route progression representations that are conserved across species. By using features derived from the non-human animal literature, we are able to test how well results generalize to the human brain. Doing so offers an insight to the validity of the various animal models for human navigation. This experiment represents the first step in closing this human-animal gap, and there remains many hypotheses to be tested in human subjects.

The route progression features used here form a topological representation of space. All routes, regardless of length and location, are normalized to the same features. This normalization preserves the sequence of locations along a route, but not their exact spatial positions. These results suggest that the human brain represents complex routes topologically rather than metrically. Note that spatial progression along a route is necessarily correlated with temporal progression along a route. To disambiguate between the two, we included both spatial and temporal route progression features in the overall banded ridge model (see section 3.2.9). Both spatial and temporal route progression models explained activity in the cortex, suggesting that there may be distinct spatial and temporal progression representations in the human brain. Here, we focused on the spatial model because of its similarities to models used in rodent electrophysiology studies.

This topological representation may confer significant computational advantages. It reduces the amount of information that needs to be represented at any time, as precise position is not needed. Doing so simplifies the representation of parts of the route that extend beyond the currently visible environment. Then, metric representations could be retrieved and updated

only as needed for the subject's current location. Furthermore, this topologic representation allows the same computational circuits to be reused for any arbitrary route, further reducing computational demands.

These results reveal that RSC, unlike OPA or PPA, is strongly selective for the start of routes. We speculate that this preference may be related to planning. RSC is implicated in reference frame transformations, and has strong anatomical projection to subcortical regions that represent allocentric cognitive maps (see section 1.2.5). At the beginning of a trial, subjects must not only recognize where they are, but also determine their position and orientation in the world, their relative position to the destination, and plan a route to the destination. These processes at the beginning of trials require both egocentric sensory representations and allocentric cognitive map representations. Because of its role in mapping between reference frames, RSC is thus likely to be heavily engaged. On the other hand, OPA is more responsible for sensorimotor transformations, and PPA is more responsible for pure scene recognition. These processes are likely to be engaged more uniformly across the entire route, and thus these two ROIs do not display the same selectivity for the start of scenes.

We note that these models did not fully recapitulate the tuning profiles of neurons found in rodent RSC and PPC. This difference may be due to several issues. First, BOLD activity in each voxel reflects the activity of hundreds of thousands of neurons. While individual neurons may be selective for different parts of a route, their overall statistics may be biased towards the start and the end of routes. Second, the taxi driver task is not identical to the rodent task. In the rodent electrophysiology studies, animals ran along pre-determined tracks. In this experiment, subjects determined their own routes. This task difference may have also elicited representations (see chapter 4). We had performed a pilot study in which subjects drove along a predetermined circuit within the map. However, this task failed to elicit similar representations of route progression in the pilot subject. It is likely that human navigation processes are engaged differently than in rodents. Finally, the parameterization of the feature spaces may be suboptimal. In this study, we divided the route into up to 16 bins. It is also possible to model route progression using a Fourier basis of sinusoids that sweep through multiple revolutions across the normalized route. We had created a set of such features, and also fit models with these features. The models perform similarly, but we found the weights to be harder to interpret than the bin-based features. Nevertheless, these results demonstrate that the human brain represents route progression information with a topological code similar to that used in rodents, and is a significant step in bridging animal electrophysiology and human neuroimaging studies of navigation.

6 Voxel-based state space modeling recovers task-related cognitive states in naturalistic fMRI experiments

This chapter is adapted from Zhang, T., Gao, J. S., Çukur, T., & Gallant, J. L. (2021). Voxel-based state space modeling recovers task-related cognitive states in naturalistic fMRI experiments. *Frontiers in neuroscience*, 14, 1103.

6.1. Introduction

To maximize efficiency and statistical power, most neuroimaging experiments use simple parametric designs and highly focused data analysis. However, the results of such experiments often do not generalize to the real world (Krakauer et al., 2017; Matusz et al., 2019; MC-K Wu et al., 2006; Yarkoni and Westfall, 2017). To address this problem, some neuroimaging experiments use more complex naturalistic conditions, such as watching movies (Hasson et al., 2004; Huth et al., 2012; Nishimoto and Gallant, 2011), listening to stories (Huth et al., 2016), or playing video games (Mathiak et al., 2011; Mathiak and Weber, 2006; Spiers and Maguire, 2007b). Neuroimaging data collected under naturalistic conditions elicit complex, dynamic patterns of brain activity across multiple functional networks that reflect the explicit and implicit task structure of the experiment (Çukur et al., 2013). Therefore, such experiments dramatically increase the complexity of data analysis and modeling. For example, while watching movies selective attention to one target or another may change the representation of information in relevant functional networks (Çukur et al., 2013), and while playing a video game the dynamic evolution of goals and subgoals over the

course of the game might evoke activity in distinct functional networks over time (Spiers and Maguire, 2006). Current methods for describing functional networks rely overwhelmingly on functional connectivity (FC) (Friston, 1994). FC analysis has become popular for use in resting state studies, where the factors driving brain activity are unknown latent variables (Blumensath et al., 2013; Buckner et al., 2013; Gordon et al., 2014; Shen et al., 2010; Smith et al., 2013; Van Dijk et al., 2010; Yeo et al., 2011). However, networks recovered by FC have no clear functional assignment. Therefore, there is a need for new methods of analysis that can efficiently describe whole-brain activity under complex naturalistic conditions.

Previous neurophysiology studies suggest that the population activity vector of a neural population represents task-related information, and that this information can be recovered even when the activities of individual neurons cannot be well-explained (Fitzpatrick et al., 1997; Pillow et al., 2008; Pouget et al., 2000; Reich et al., 2001). Indeed, many neurophysiology studies have treated the activity of multiple neurons as a dynamical system whose state represents task variables. For example, in arm reach tasks, motor neurons in the non-human primate brain traverse uniquely identifiable trajectories in their state space during reaches to distinct targets (Srinivasan et al., 2006; W Wu et al., 2006). In decision making tasks, the state of prefrontal neurons in non-human primates represents several task-relevant variables including the attentional target (Mante et al., 2013). In navigation tasks, the state of parietal neurons in rats represent information about upcoming turns and planned routes (Harvey et al., 2012). In a goal-directed navigation task, the state of neurons in the retrosplenial cortex (RSC) in rats represents cues and reward locations (Vedder et al., 2017). In both rats and non-human primates, the state of a population of neurons also encodes information about a choice that the animal made (Harvey et al., 2012; Mante et al., 2013; Vedder et al., 2017).

These studies provide strong evidence, from both non-human primates and rodents, that task variables are systematically embedded in a task-related state space that is distributed across neural populations. The dimensionality of this task-related state space appears to be markedly lower than the dimensionality of the total activity space spanned by the population activity (Harvey et al., 2012; Mante et al., 2013; Vedder et al., 2017). For example, in a rat navigation task, a 3-dimensional subspace of the total activity space of 65 neurons is sufficient to distinguish choices made by the animal (Harvey et al., 2012). In non-human primates, task variables and the animals' decisions are encoded a 4-dimension subspace of hundreds of neurons (Mante et al., 2013). It is thus likely that the human brain also represents task variables in a low-dimensional subspace of its total activity space. Furthermore, state-space approaches may be particularly useful for understanding brain function in complex, naturalistic tasks that activate complex networks of brain areas (Huth et al., 2016; Spiers and Maguire, 2007b). Therefore, analyzing data in the context of the population activity vector may offer new insights about task-related representations in the human brain.

Here we develop a voxel-based state space modeling method for analyzing fMRI data under naturalistic conditions. Our framework is inspired by methods developed originally to model primate electrophysiology data (Mante et al., 2013). The framework is based on the idea that task variables, such as stimulus information and cognitive states of the subject, are

represented implicitly in the population activity vector of the entire cortex. The framework specifies a simple algorithm that finds the subspace of the entire activity space that best represents these task variables. We use this framework to successfully recover and interpret task-related state spaces in two naturalistic fMRI experiments: a controlled visual attention task and an open-ended video game.

6.2. Methods

6.2.1. Experimental paradigm

6.2.1.1. Subjects

Seven healthy volunteers (six males, one female; ages 25-32) with normal or corrected to normal vision participated in the experiments. Six subjects (S1-S6) participated in the visual search experiment that was published as a part of a previous study (Çukur et al., 2013). Two male volunteers (age 25 & 26) participated in the second video game experiment that was run as a part of a pilot experiment for a separate purpose (subjects S1 & S7). The experimental procedures were approved by the Institutional Review Board at the University of California, Berkeley. Written informed consent was obtained from all subjects.

6.2.1.2. Experiment procedure

Subjects participated in two experiments, a visual attention task and a first-person shooter video game. We used the visual attention task to investigate whether representations of known task variables in a relatively controlled naturalistic task could be recovered from the population activity vector of the brain. We used the video game to investigate whether representations of task variables could be recovered from brain activity during a complex and open-ended task, and also whether a data-driven method might recover these representations.

In the visual attention task (Çukur et al., 2013), subjects viewed short natural movies while steadily fixating on a central dot. In separate runs, subjects covertly attended either to the presence of humans or the presence of vehicles in the movies. Humans and vehicles appeared in a diverse array of settings and in many different sizes, positions, and orientations. In some frames, both humans and vehicles were present. Subjects were instructed to respond with a button press when an exemplar of the attended category was present on the screen. Data for each attention condition were collected across three 10-minute runs. In separate sessions, subjects passively viewed an extended set of natural movies without performing visual search. All attention conditions were mutually exclusive (i.e. there were no “attend to humans and vehicles” condition). Here we analyzed data from five subjects included in the original visual search study (Çukur et al., 2013) as well as one additional subject whose data was acquired subsequently.

In the video game task, subjects engaged in a simulated first-person combat against computer players in Counter Strike: Source (Valve Co. Bellevue, WA). In this game the subject played a member of a counter-terrorism force attempting to stop a terrorist force from planting a

bomb. The first subject controlled the game using two button fiber optics response pads and an MR-compatible trackwheel mouse (Current Designs, Philadelphia, PA), while the second subject used a gamepad that was modified to be MR-compatible. To alleviate the difficulty of gameplay in the scanner and to better capture game dynamics, gameplay speed was slowed down to 50% of the normal speed. Because Counter Strike: Source is an interactive open-ended game, it was not possible to obtain identical repeats of the audiovisual stimuli or player actions. Gameplay was recorded using a system built into the game engine. After the MR experiment, the audiovisual stimulus during the gameplay was reconstructed based on these recorded data. Ninety minutes of data was collected across six 15-minute runs for one subject and forty-five minutes of data was collected across three 15-minute runs for the second subject. Because these data were originally collected as a pilot project, the subjects differed in the controls used and amount of time spent in the task.

6.2.1.3. Task Variables

The supervised state space method requires explicit definition of task variables. We therefore operationally defined task variables for each of the two tasks. Here, "task variables" can encompass both task-relevant stimulus features, and also states endogenous to the subject such as attention.

6.2.1.3.1. Visual attention task.

The relevant task variables are the attentional state of the subject (attend to humans, attend to vehicles, or passive), the presence of humans in the movies, and the presence vehicles of the movies. Because the search for humans and for vehicles were performed in distinct runs, the attentional conditions were mutually exclusive. Thus, they are grouped under a single task variable. Taken together, these three variables define a three-dimensional state space that encompasses both stimulus- and task-related information.

6.2.1.3.2. Video game task.

Because the video game task is complex and open-ended, ground-truth task variables are not available. To obtain reliable, human-interpretable task variables, two human observers inspected the game recordings and together they defined a set of nine behavioral states for the video game task: "dead", "round start", "safe explore", "unsafe explore", "kill", "engage close", "engage far", "flash", and "run away." (The full definition of each state is given in Table 6.1.) The observers then labeled each TR with these states. The states were not mutually exclusive. For this analysis, each behavioral state was treated as a task variable. Taken together, these variables define a nine-dimensional task space for the video game task that captures the behavioral state of the subject.

6.2.2. MRI Protocols

Functional MRI data were collected using a 32-channel head coil on a 3T Siemens Tim Trio scanner at the University of California, Berkeley. Functional data were acquired using a T2*-weighted gradient-echo EPI sequence customized with a water-excitation radiofrequency pulse to prevent contamination from fat signal. The following parameters were prescribed:

Table 6.1 Descriptions of the nine behavioral states in the video game task. Because ground-truth task variables were not available in the open-ended video game task, two human observers inspected the game recordings and defined nine behavioral states. These states capture the range of behaviors of the subjects in the video game task and serve as human-interpretable task variables. These states were used to label each TR for the state space analyses. The states and their definitions are given here.

Behavioral State	Definition
Dead	The subject's character is dead and the subject is unable to participate in the game
Round start	Prep time at the beginning of a round
Safe explore	Subject is exploring the map when it is known that the enemy is too far away to engage
Unsafe explore	Subject is exploring the map but there is a possibility of engaging enemies
Kill	Subject has killed an enemy
Engage close	Subject is engaging enemies in close quarters combat
Engage far	Subject is engaging enemies at a distance
Flash	Subject is hit with a flashbang and is unable to see
Run away	Subject is running away from enemies

repetition time = 2 s, echo time = 34 ms, flip angle = 74°, voxel size = 2.24 × 2.24 × 3.5 mm^3, field of view = 224 × 224 mm^2, matrix size = 100 × 100, and 32 axial slices to cover the entire cortex. Data were acquired in 325-volume runs. The first and last 10 volumes of each run were discarded. No acceleration was used. Head motion was minimized with foam padding. To enable reconstruction of cortical surfaces, anatomical data were acquired using a three-dimensional T1-weighted MP-RAGE sequence with the following parameters: voxel size = 1 × 1 × 1 mm^3 and field of view = 256 × 212 × 256 mm^3.

6.2.3. Preprocessing and Visualization

Preprocessing of functional data was performed using an in-house processing pipeline. For each subject, the Statistical Parameter Mapping toolbox (SPM8, http://www.fil.ion.ucl.ac.uk/spm/software/spm8/) was used to align all brain volumes to the first volume in the first run. Non-brain tissues were excluded using the Brain Extraction Tool (BET, http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/BET). Cortical surfaces were reconstructed using Freesurfer (Dale et al., 1999). All analyses were performed on cortical voxels. All stages of preprocessing were checked carefully by hand.

Low-frequency drifts in single voxel responses were estimated using a 240-s Savistsky-Golay filter of third order and removed from the responses. The detrended voxel responses were normalized to zero mean and unit variance. No spatial smoothing or filtering was used. Functional regions of interest (ROIs) in each subject were localized using functional localizer and retinotopic mapping data collected for this purpose. Boundaries of functional ROIs were delineated on each subject's cortex based on relative response levels to standard functional localizers (Huth et al., 2012). Mapping of functional signals to the cortical surface and visualization of results on the cortical surfaces was performed using the pyCortex toolbox (Gao et al., 2015).

It is possible that subcortical structures and the cerebellum are activated during these naturalistic tasks, particularly the video game task that requires motor commands. However, the pulse sequence used in this study was optimized for cortical signals and the slice prescription did not include the cerebellum. Therefore, the data were insufficient to permit modeling of subcortical voxels and the cerebellum, and so all analysis procedures were restricted to cortical voxels.

6.2.4. Deconvolution of the hemodynamic response in the attention task

As a further step in the preprocessing, we deconvolved the hemodynamic response function (HRF) from data in the attention task. BOLD measurements reflect delayed hemodynamic responses consequent to underlying neural activity (Kay, David, et al., 2008; Ogawa et al., 1990). The HRF lasts many seconds, peaking at 5-10 seconds from the onset (Glover, 1999). In the visual attention task, the presence of target categories can switch rapidly on the order of seconds. The HRF effectively smooths the measured brain activity and causes states between successive TRs to be less distinguishable from each other than they might be otherwise. This smoothing could hamper state space analyses. Therefore, we aimed to deconvolve the HRF from the raw brain activity by fitting a voxelwise encoding model. Briefly, voxelwise modeling (VM) (Kay, Naselaris, et al., 2008; Naselaris et al., 2009, 2011; Nishimoto et al., 2011) treats the activity of each voxel as a linear transformation of stimulus features. The HRF is modeled by fitting a separate finite impulse response (FIR) filter to every distinct feature separately for every voxel. The FIR filter is implemented by concatenating feature vectors that had been delayed by 1, 2, 3, and 4 TRs. In the attention task, subjects focused on the semantic contents of the stimulus. Therefore, we use the WordNet semantic category labels as used in our earlier study (Çukur et al., 2013) as a basis for the HRF deconvolution. This model was previously shown to provide comprehensive descriptions of the responses of visual and non-visual cortical voxels to semantic information (Çukur et al., 2013, 2016; Huth et al., 2016). To construct timecourses of the WordNet features, salient semantic categories in each 1-sec movie clip were manually labeled, and then temporally downsampled to match the fMRI sampling rate. Ridge regression was used to estimate weights with 10-fold cross validation. For each voxel, the set of weights across the FIR were averaged to obtain a single weight per feature. These average weights reflect the mean selectivity of each voxel to each feature without temporal smoothing. They were then multiplied with the feature timecourses to estimate brain activity deconvolved from the HRF. This deconvolution procedure greatly improved the interpretability of state space results on the attention task.

In this study, we used existing semantic labels from previous experiments for the visual attention data. However, no such labels existed for the video game data, and only behavioral labels were available. Furthermore, it is challenging to compile an explicit feature space a priori that captures all aspects of the video game task. Therefore, no HRF deconvolution was applied in the video game task, and state space modeling was performed directly on

measured BOLD responses. Note that, as a result, the recovered states might be less separable in the state space than if the responses had been deconvolved using an appropriate feature space.

6.2.5. Voxel-based State Space Modeling

To recover task-related state spaces from preprocessed BOLD responses, we adopt targeted dimensionality reduction, a modeling framework originally devised for primate electrophysiology experiments (Mante et al., 2013). This framework assumes that task variables are represented in the activity vector of all cortical voxels. The task-related state space is a subspace within the entire space spanned by the cortical activity vector. This subspace captures the variance in cortical activity that can be attributed to the representation of task variables. The dimensions of this task-related state space collectively reflect the representation of task variables. Each task-related brain state can then be characterized by a unique set of state variable values. The projection of the cortical activity vector into this state space characterizes the representation of state variables in the brain at each point in time. In the current paper we make no assumptions about specific brain regions that may be involved in representation of the task-related state space. Thus, all cortical voxels (voxels that fall between the pial and white matter surfaces) are analyzed in each subject (48,673 ± 4,382 voxels, mean ± std across subjects). However, this same analysis could easily be applied to specific regions of interest (ROIs) by analyzing only the activity of voxels within those ROIs.

The voxel-based state space method consists of two steps. First, a set of task variables that are assumed to capture the underlying structure of the task are operationally defined (see section Task Variables). Then a low-dimensional task-related state space that is hypothesized to represent these task variables is recovered by regressing task variables on to cortical activity.

Here, cortical activity $\boldsymbol{Y}$ was modeled as a linear combination of the task variables $\boldsymbol{X}$. The timecourses were regressed against all task variable timecourses (three for the visual attention task, and nine for the video game task) using ordinary least squares (OLS) regression. We use OLS instead of ridge regression here because the number of state variables is far fewer than the number of samples, and thus overfitting is not likely to be a concern. This step is identical to traditional univariate analyses of voxelwise regression with binary categorical features. This procedure yielded a set of weights $\boldsymbol{B}$ that map task variables to cortical activity (size [variables × voxels]) ($\boldsymbol{Y} = \boldsymbol{XB}$).

The set of weights obtained for each task variable forms a vector of length [voxels] in the cortical activity space. Each weight vector represents the characteristic pattern of cortical activity associated with the representation of that task variable. However, these weights are inherently noisy since they can implicitly capture other non-task-related activity patterns irrelevant to this experiment and which account for little variance. To focus on the task-driven activity, a denoiser was created using principal component analysis (PCA) as was performed in the original electrophysiology analysis method (Mante et al., 2013). PCA was performed on the timecourse $\boldsymbol{Y}$ of cortical activity (size [TRs × voxels]) in individual subjects. The first 24 eigenvectors of $\boldsymbol{Y}$, $\boldsymbol{U}_L$, were retained (size [voxels × 24]), accounting for ~70% of the total variance in BOLD responses, and a denoiser $\boldsymbol{D}$ was calculated as $\boldsymbol{D}=\boldsymbol{U}_L\boldsymbol{U}_L^T$.

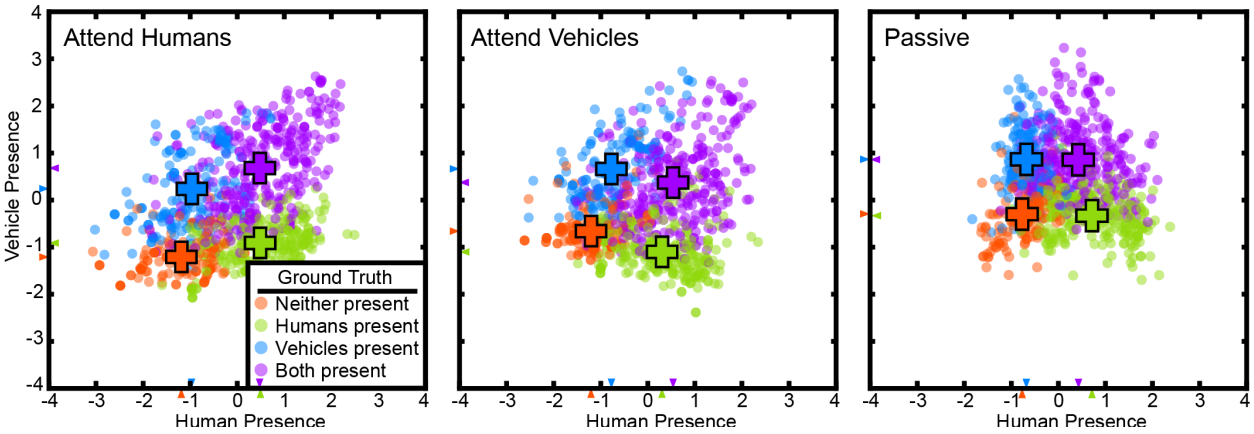


Figure 6.1 The state space method recovered a task-related state space in the visual attention task. Cortical activity at each TR for subject S2 is projected to a point in this space. The state space is projected on to the 2D plane spanned by the "human presence" (horizontal) and "vehicle presence" (vertical) axes. Projections are shown for the attend-human, attend-vehicles, and passive viewing conditions. Positive values indicate presence, and negative values indicate absence. TRs are color-coded by their ground truth states. Orange: neither present; blue: only humans present; green: only vehicles present; purple: both humans and vehicles present. Crosses indicate the mean positions of each group. The cluster separation index is 0.75 for attend-humans, 0.69 for attend-vehicles, and 0.67 for passive ($p < 1e\text{-}5$, permutation test), all indicating significant separation of the clusters. Clusters are more distinct in the attentive conditions than in the passive condition ($p < 1e\text{-}5$, permutation test). These results are consistent with the hypothesis that task variables are represented in a low-dimensional task-related subspace of the cortical activity space, and that attention increases the separation of the states in this task-related state space.

Next, $\boldsymbol{B}$ was denoised using $\boldsymbol{D}$ to create $\boldsymbol{B}_L=\boldsymbol{DB}$ and to form the state space. Note that the weight vectors of the task variables are not guaranteed to be orthogonal. To form an orthonormal basis for the task-related state space, QR decomposition was performed on the transpose of $\boldsymbol{B}_L$ (size [variables × voxels]), factorizing $\boldsymbol{B}_L^T=\boldsymbol{QR}$. The first [variables] columns of the resulting $\boldsymbol{Q}$ (size [voxels × voxels]) matrix, $\boldsymbol{Q}_S$, then represent the dimensions of the state space. Each dimension in this state space reflects the brain's representation of one task variable. Finally, the cortical activity vector at each TR was projected to a point in this state space ($\boldsymbol{p}_i = \boldsymbol{y}_i\boldsymbol{Q}_S$). This point is the estimate of the state of the brain at that TR.

6.2.6. Quantification of state space results

We developed two convergent metrics to validate the performance of the state space model. The first metric is based on the assumption that a good state space should allow clear separation of different states. To quantify this we defined a cluster separation index (CSI) as follows. First, a multivariate normal distribution was fitted to each cluster. Next, all pairwise Jensen-Shannon divergences (JSD) between the clusters were calculated (Lin, 1991). JSD provides a measure of the difference between two distributions by averaging the Kullback-Leibler divergence (KLD) (Kullback and Leibler, 1951) of individual distributions from their average distribution. We chose the JSD because it is symmetric and has a finite upper bound. Finally, the CSI was taken as the average of all inter-cluster JSDs. CSI ranges from 0 to 1,

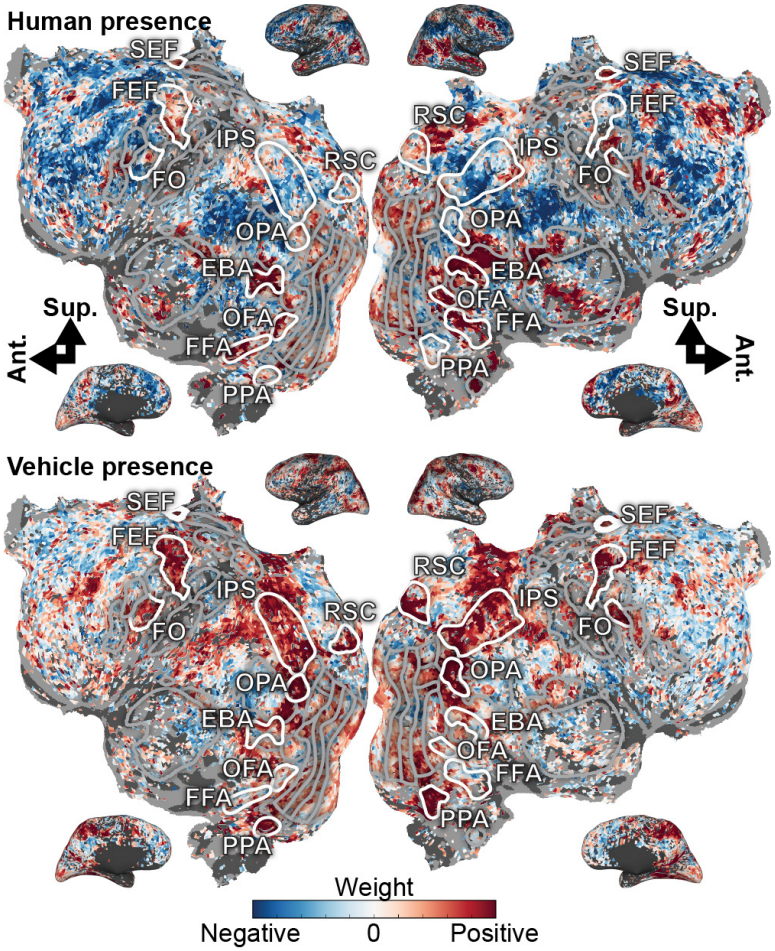


Figure 6.2 Task variables in the visual attention task are represented in broad functionally specialized networks distributed across the cerebral cortex. Model weights for human- and vehicle-presence are shown on the flattened cortical surface for subject S1. Blue correspond to negative weights, white to no weights, and red to positive weights. Voxels are thresholded by the prediction performance of a semantic encoding model at a non-FDR-corrected $p < 0.05$ significance level. Selected ROIs are highlighted and labeled. Inflated cortices are shown for reference. The human presence axis is represented in FFA, OFA, and EBA, which are known to be activated by faces or body parts. Additionally, it is represented in MT, LO, SEF, and around the inferior frontal sulcus. The vehicle presence axis is represented in RSC, OPA, and PPA, all activated during spatial perception or navigation; in the IPS and FEF, both involved in spatial attention; and also in FO and inferior frontal sulcus. These model weights agree well with the known functions of the cortex and further suggest that task variables are meaningfully represented in a low-dimensional subspace of cortical activity.

where 0 indicates no separation and 1 indicates complete separation. While visualization occurs in 2D space, CSI values were calculated in the original state space (3 dimensions for the visual attention task and 9 dimensions for the video game task).

The second metric is based on the assumption that any valid state space model should accurately predict state given new cortical activity. To quantify this we used a 10-fold cross-validation procedure to evaluate state-space prediction performance in each individual subject. In each fold, 10% of the data were held out as the validation set, and the proposed method was applied on the remaining portion to learn the state space. Each TR in the

training set was labeled using the same labels used for the supervised method. The visual attention task used 12 labels (the four labels for human and vehicle presence × three attention conditions), while the video game task used nine labels (the set of behavioral states). Afterwards, each TR in the validation set was projected into the learned space, and assigned to the nearest cluster to classify its state. This label was then compared against the ground truth label. Classification accuracy was averaged across cross-validation folds. For both metrics, significance was evaluated using permutation tests. Null models were created by permuting the state labels, and performing the full analysis using the permuted labels.

6.3. Results

To understand how task information may be represented in the population activity space of the cortex, we developed a voxel-based state space framework for fMRI. The task-related state space is recovered by regressing task variables directly on to cortical activity, and activity at each TR can then be projected to a point in this task-related state space. We use this framework to recover low-dimensional task-related state spaces from cortical activity recorded during a visual attention task and a video game task.

6.3.1. The voxel-based state space method recovers a state space for variables in the attention task

In the attention task, subjects watched short, naturalistic movies while attending to the presence of humans or vehicles in different runs, or they watched the movies passively. The visual attention task had three main variables: the presence of humans, the presence of vehicles, and the attention target of the subject. We applied the voxel-based state space method to data from this task to examine whether task variables are represented in the cortical activity vector. The voxel-based state space method was used to learn a state space on data from all attentional conditions for each subject separately. Figure 6.1 shows the recovered state space for one subject, projected onto the 2D plane defined by human- and vehicle-presence axes (for all subjects see Supplementary Figure 9). On this plane, TRs were labeled with the states of "only humans present," "only vehicles present," "neither present," and "both present."

Animal studies show neural activities corresponding to different choices projected to distinct regions in their state space (Harvey et al., 2012; Mante et al., 2013; Vedder et al., 2017). Based on this result we hypothesize that a good state space should allow clear separation of different states. Therefore, the state space was evaluated by measuring how well these states were separated. A cluster separation index (CSI) was defined as the average pairwise Jensen-Shannon divergences of the state clusters. The states are distinctly separated with a CSI of 0.66 ± 0.09 (mean ± std across subjects) for attend-humans, 0.74 ± 0.03 for attend-vehicles, and 0.53 ± 0.24 for passive viewing ($p < 1e\text{-}5$; permutation test). These results show that task variables are represented in a low-dimensional task-related state space that can be recovered successfully by the supervised state space method. Furthermore, we found that the CSI in the passive condition was significantly lower than the CSI in the attentive conditions for five out of the six subjects ($p < 1e\text{-}5$; permutation test).

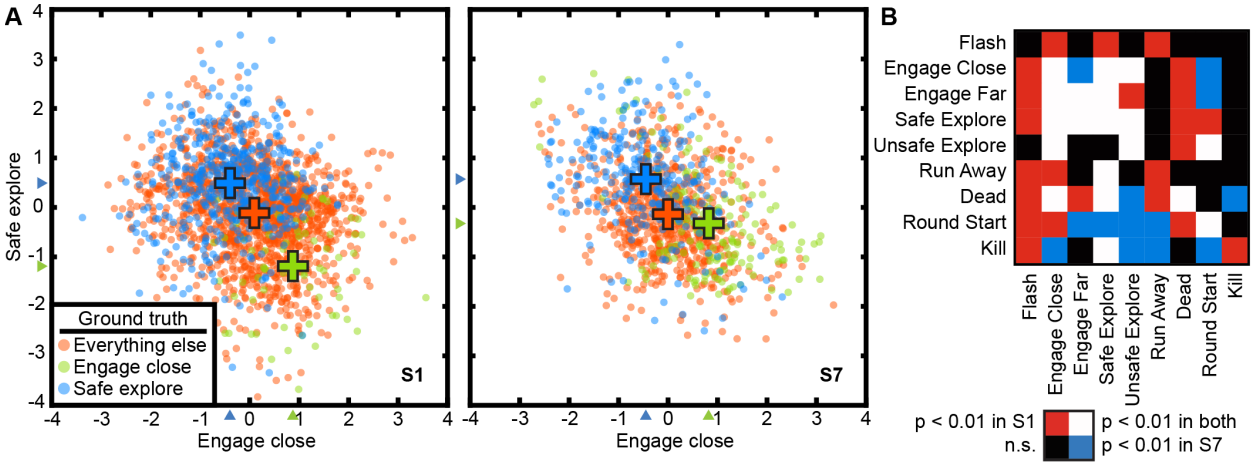


Figure 6.3 The state space method recovered a task-related state space in the open-ended video game task. Cortical activity at each TR is projected to a point in a 9-dimensional state space. The state space method revealed significant relationships between the nine behavioral states in the video game task. The nine states are not orthogonal, and some states are likely to lead to others. (A) The state space projected to two dimensions for "safe explore" and "engage close" for both subjects. TRs are color-coded by their ground truth state. Blue: safe explore; Green: engage close; Orange: all other states. Crosses indicate the mean positions of each group. The CSI was 0.72 for subject C1 and 0.62 for subject C2, indicating significant separation of the clusters ($p < 1e\text{-}5$, permutation test). These two states are mutually exclusive and score negatively on each other's axis. (B) Relationships between states are reflected in non-zero projections of states on to other state dimensions. Rows and columns are behavioral states as described in Table 1. The table is asymmetric because relationships between states are not necessarily reciprocal. Colors indicate Bonferroni-corrected significance ($p < 0.01$, permutation test). Red: significant for subject C1; Blue: significant for subject C2; White: significant in both subjects; Black: not significant in either subject. Because the task is open-ended, subject may adopt different strategies and thus the exact pattern of correlation differs between subjects. Task variables in a complex, open-ended task are represented in a low-dimensional space in cortical activity that also captures relationships between states.

Any valid state space model should accurately predict brain states in a new data set that was not used to fit the model. To test this, cross-validated (10-fold) classification performance in each subject was evaluated using 12 distinct labels (the four labels for human and vehicle presence × three attention conditions). Average performance across subjects is 48.4 ± 7.6% ($p < 1e\text{-}5$, permutation test, chance = 8.3%). This result indicates that the recovered state space can be used to predict cortical state in a separate test data set.

To better understand the cortical representations of the recovered state space, we projected the weights for each dimension of the task-related subspace onto the cortical surface (Figure 6.2 and Supplementary Figure 10). The "human presence" dimension projects heavily onto FFA (Kanwisher et al., 1997), OFA (Gauthier et al., 2000), and EBA (Downing et al., 2001), functional areas that represent faces and/or body parts. This dimension also projects onto MT+ (Tootell et al., 1995; Zeki et al., 1991), which represents motion and human movement, LO (Grill-Spector et al., 1999), which is activated during object perception, and the frontal operculum (FO) (Corbetta et al., 1998), activated in spatial attention tasks. The "human presence" dimension also projects to the inferior frontal sulcus, which is known to contain a face-selective region (Avidan et al., 2005; Tsao and Livingstone, 2008). In contrast, the

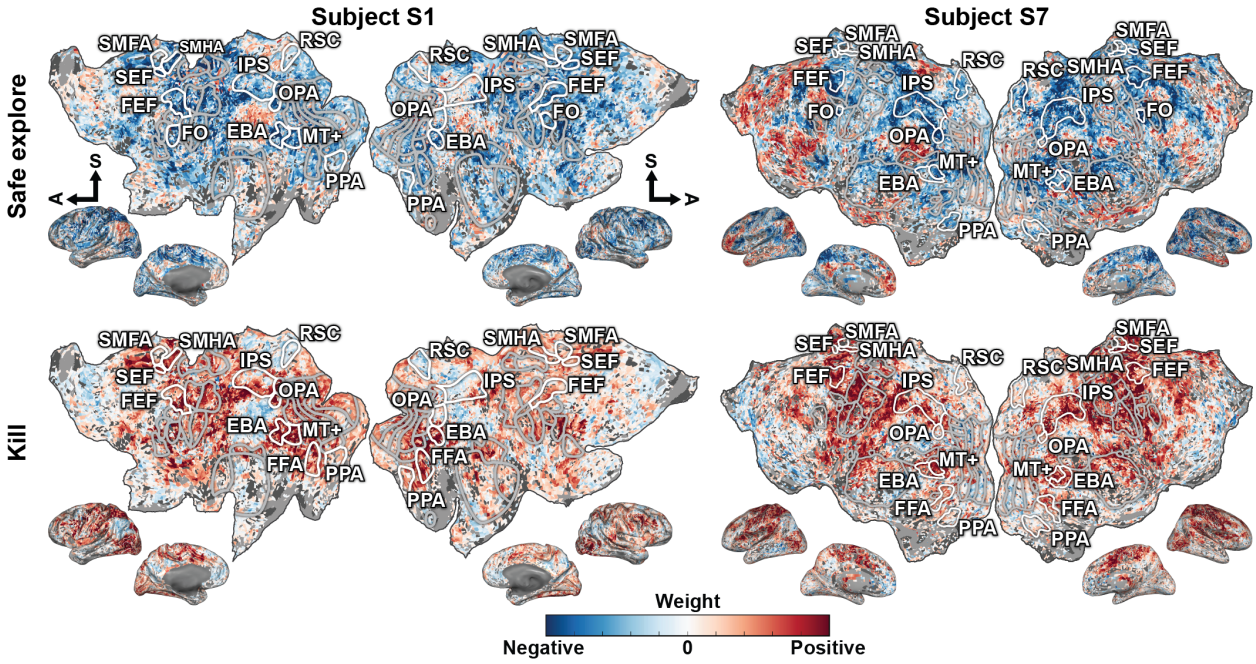


Figure 6.4 Behavioral states in the video game task are represented in broad functionally specialized networks distributed across the cerebral cortex. The representations fall into two general groups that share overlapping regions. Weights for one state dimension from each group, "safe explore" and "kill", are shown on the flattened cortical surfaces for both subjects. Blue correspond to negative weights, white to no weights, and red to positive weights. Voxels are thresholded by the prediction performance of an encoding model at a Bonferroni-corrected $p < 0.05$ significance level. Selected ROIs are highlighted and labeled. Inflated cortices are shown for reference. The axes for "dead," "safe explore," "round start," "flash," and "unsafe explore" share representation in the TPJ, precuneus, and prefrontal cortex. The axes for "kill," "engage close," "run away," and "engage far" share representation in the motor, pre-motor, and supplementary motors areas, and also in IPS, FEF, and SEF. These model weights further suggest that task variables in complex and open-ended tasks are also represented in a low-dimensional subspace of cortical activity.

"vehicle presence" dimension projects strongly onto RSC (Epstein, 2008; Maguire, 2001), PPA (Epstein et al., 1999), and OPA (Dilks et al., 2013), regions known to be active during spatial perception or navigation. The "vehicle presence" dimension also projects onto the IPS (Kastner et al., 1999; Silver et al., 2005) and frontal eye field (FEF) (Corbetta et al., 1998; Kastner et al., 1999; Moore and Fallah, 2001; Paus, 1996), both of which are activated in tasks requiring spatial attention. Finally, the "vehicle presence" dimension projects around the inferior frontal sulcus. These results show that the dimensions of the task-related state space recovered by the voxel-based state space method are related systematically to the functional selectivity of the brain regions that are activated under each state.

6.3.2. The voxel-based state space method recovers a behavioral state space for the open-ended video game task

Because the visual attention task involves explicit task states, it is relatively straightforward to perform a state space analysis of those data. It is more difficult to perform a state space analysis of complex, open-ended tasks. To explore this issue, we applied the supervised

method to a video game task. In this task, subjects engaged in simulated combat in Counter Strike: Source. Each TR was labeled using nine experimenter-defined behavioral states (Table 6.1). These were used as task variables for the supervised method. Because it is impractical to fully decorrelate the multitude of states that arise during complex naturalistic tasks, these states are partially correlated. Thus, this analysis also serves to test whether the state space method can recover a task-related state space even when states are correlated.

A 9-dimensional task-related state space was learned for the video game task for each subject using the voxel-based state space method. Figure 3 shows two dimensions of the recovered state space for both subjects, projected onto two planes each defined by "engage close" and "safe explore." These two dimensions were chosen as the corresponding brain states project not only on to their own dimension, but also on to the other. States are labeled using the nine behavioral labels shown in Table 6.1. The states are distinctly separated with a CSI of 0.72 in subject S1 and 0.62 in subject S7 ($p < 1e\text{-}5$, permutation test). Thus, task variables appear to be represented in a low-dimensional space even in this complex, open-ended task.

To gain a better understanding of how state correlations affect these low-dimensional task representations in brain activity, we examined the projection of states on to other dimensions in the task-related state space. We find that when states are correlated, TRs associated with one state project to a non-zero position in the dimension associated with the other state. For example, "engaging close" and "safe explore" each project negatively to the dimension associated with the other state (Figure 6.3A) ($p < 1e\text{-}5$, permutation test). The full list of significant correlations between states for the two subjects is given in Figure 6.3B. The exact patterns of correlations differ between subjects; because the video game task is open-ended, the subjects used different strategies and these resulted in different game dynamics and different correlations. Nonetheless, the correlations between states are captured by the low-dimensional representations in the brain activity of both subjects.

As noted earlier, any valid state space model should accurately predict brain states in a new data set that was not used to fit the model. We therefore performed a separate test to determine whether the state space can be used to predict cortical state in the video game task. Because the video game task is open-ended, the behavioral states are not all equally like to occur. Thus, a simple accuracy score may misrepresent prediction accuracy by, for example, simply always guessing the most common class. We therefore used the balanced prediction accuracy metric (Brodersen et al., 2010), which accounts for the unequal frequency of classes by averaging the recall score on each class. Average balanced accuracy across 10 cross-validation folds in the subject is 32.8 ± 3.8% in subject S1 and 30.4 ± 4.2% in subject S7 ($p < 1e\text{-}5$, permutation test, chance = 11%). This result suggests that the learned state space can be used to predict cortical state even in the presence of correlations between states.

To understand how the recovered states are represented in the brain, the model weights for each dimension were projected onto the cortical surface for both subjects (Figure 6.4 shows two dimensions for both subjects, for all dimensions see Supplementary Figure 11 and 12). They appear to fall into two general categories. The dimensions for "safe explore," "round start," "flash," "dead," and "unsafe explore" project on to the TPJ, precuneus, and prefrontal cortex. The dimensions for "run away," "engage close," "kill," and "engage far" project on

to the motor, pre-motor, and supplementary motors areas, and also in IPS, FEF, and SEF. Because the specific task states examined here have not been investigated in previous studies, it is difficult to relate these functional assignments to neuroimaging literature. However, it appears that the first category predominantly engages areas within the default-mode network, while the second category predominantly engages areas that represent selective attention and motor movement. The weights for each dimension are largely consistent across the two subjects. These results suggest that there is a low-dimension representation of task variables in cortical activity even in more complex and open-ended tasks.

6.4. Discussion

Complex naturalistic behaviors evoke complex, high-dimensional activity across the cortex that are challenging to analyze and interpret (Çukur et al., 2013; Hasson et al., 2004; Huth et al., 2012; Mathiak et al., 2011; Mathiak and Weber, 2006; Nishimoto et al., 2011; Spiers and Maguire, 2007b). To understand this complex activity, we developed a voxel-base state space modeling framework to find interpretable, low-dimensional representations of task variables. In both a visual attention task and a video game task, we find a low-dimensional representation of task variables in cortical activity. Furthermore, the cortical areas associated with representations during the attention task agree well with previous studies.

These low-dimensional representations also capture both attentional effects on representation in the attention task and correlations between representations in the video game task. In previous studies (Çukur et al., 2013; David et al., 2008) we found that attention warps semantic representations by increasing the distance between categories in semantic space. Here, the CSI in the passive condition was significantly lower than the CSI in the attentive conditions in all but one subject. Thus, attention to semantic categories significantly increases the separation of states in the task-related subspace relative to what is found under passive viewing. The increased separation in state space directly reflects attentional warping: both serve to increase the cortical representations of task-relevant variables to facilitate the task. This finding agrees with other studies that also found attention improved representation of features, such as color (Brouwer and Heeger, 2013), physical and conceptual properties (Harel et al., 2014), or animal taxonomy and behavior (Nastase et al., 2017). Unlike these prior studies, here we were able to demonstrate the increased separation directly in the activity space of the brain and using much more naturalistic stimuli. Thus, these data provide further evidence that attention warps functional representations to improve task (Çukur et al., 2013; David et al., 2008). In the video game task, the behavioral states could not be counterbalanced and therefore were not independent of each other. In the task-related state space, we find that representations of correlated behavioral states are not orthogonal. Correlated states project to non-zero points on each other's dimension in state space, showing correlated task variables have similar cortical representations.

There are several limitations concerning the task-related state space we recovered from the rich video game data. First, we used nine behavioral labels for the state space data that were hand-picked by observers. While the labels were agreed upon between two observers familiar with the video game, different observers might pick different labels and thus would produce a

different state space. Second, behavioral labels can only serve as a proxy to true task variables. An objective parameterization of the video game task would produce a better set of variables from which to build the task-related state space. Using machine learning to parametrize the game state may be one such avenue for future research (Jaderberg et al., 2019). Alternatively, a completely unsupervised method may be able to recover a task-related state space in a completely data-driven manner. However, using an unsupervised method would raise the issue of interpreting the recovered state spaces in the context of the tasks. Last, we collected only a very limited data set using the video game task. Therefore, the data may have simply been too noisy to allow us to recover the optimal task-related state space from cortical data. Nevertheless, the current results support the existence of low-dimensional representations even in complex and open-ended tasks, and suggest that future work focusing on the video game task is promising.

Many other methods have been proposed for modeling high-dimensional cortical activity. Multivoxel pattern analysis (MVPA) (Cox and Savoy, 2003; Haxby et al., 2001; Kamitani and Tong, 2005; Norman et al., 2006) decodes the stimulus or task from the activity of a set of voxels to provide a categorical description of brain activity. The classifier used to evaluate the quality of the state space model is mathematically similar to MVPA in that some stimulus or task variable is decoded from the activity of a collection of voxels. However, the state space method provides a model of how task variables are continuously encoded in the cortical activity vector of the brain. The classifier is distinct from this model and does not form the core of the state space method. The classifier is not used to interpret the recovered state space, and is used only to evaluate the quality and stability of the recovered state space. A few studies (Billings et al., 2017; Shine, Breakspear, et al., 2019; Shine, Hearne, et al., 2019) have also attempted to recover low-dimensional cortical activity in data from the Human Connectome Project (HCP) (Van Essen et al., 2013). However, because the HCP used multiple tasks and collected extremely limited amount of data per task, these studies can only examine the general activity differences between tasks. They cannot describe how variables within each task are represented in cortical activity.

In sum, our results suggest that while overall cortical activity is high-dimensional, the representation of particular task variables is reflected in a low-dimensional subspace of brain activity. Analyzing the cortical activity vector of the brain complements analyses on single voxel activities by revealing the complex interactions between brain systems. These interactions can be useful for discovering functionally linked networks of brain regions. We expect the voxel-based state space method will be useful in understanding high-dimensional brain activity elicited by complex, open-ended naturalistic tasks.

7 References

Aguirre, G.K., Zarahn, E., and D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: evidence and implications. Neuron 21, 373–383.

Alexander, A.S., and Nitz, D.A. (2015). Retrosplenial cortex maps the conjunction of internal and external spaces. Nat. Neurosci. 18, 1143–1151.

Alexander, A.S., and Nitz, D.A. (2017). Spatially Periodic Activation Patterns of Retrosplenial Cortex Encode Route Sub-spaces and Distance Traveled. Curr. Biol. 27, 1551-1560.e4.

Andersen, R.A., and Buneo, C.A. (2002). Intentional Maps in Posterior Parietal Cortex. Annual Review of Neuroscience 25, 189–220.

Aronov, D., and Tank, D.W. (2014). Engagement of neural circuits underlying 2D spatial navigation in a rodent virtual reality system. Neuron 84, 442–456.

Auger, S.D., Zeidman, P., and Maguire, E.A. (2015). A central role for the retrosplenial cortex in de novo environmental learning. ELife 4.

Avidan, G., Hasson, U., Malach, R., and Behrmann, M. (2005). Detailed Exploration of Face-related Processing in Congenital Prosopagnosia: 2. Functional Neuroimaging Findings. Journal of Cognitive Neuroscience 17, 1150–1167.

Behzadi, Y., Restom, K., Liau, J., and Liu, T.T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 37, 90–101.

Bellmund, J.L., Deuker, L., Navarro Schröder, T., and Doeller, C.F. (2016). Grid-cell representations in mental simulation. Elife 5.

Bergstra, J., Yamins, D., and Cox, D. (2013). Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. In Proceedings of the 12th Python in Science Conference, (SciPy), p.

Billings, J.C.W., Medda, A., Shakil, S., Shen, X., Kashyap, A., Chen, S., Abbas, A., Zhang, X., Nezafati, M., Pan, W.-J., et al. (2017). Instantaneous brain dynamics mapped to a continuous state space. Neuroimage 162, 344–352.

Blumensath, T., Jbabdi, S., Glasser, M.F., Van Essen, D.C., Ugurbil, K., Behrens, T.E.J., and Smith, S.M. (2013). Spatially constrained hierarchical parcellation of the brain with resting-state fMRI. Neuroimage 76, 313–324.

Bonner, M.F., and Epstein, R.A. (2017). Coding of navigational affordances in the human visual system. Proc. Natl. Acad. Sci. U. S. A. 114, 4793–4798.

Brainard, D.H. (1997). The Psychophysics Toolbox. Spat. Vis. 10, 433–436.

Brodersen, K.H., Ong, C.S., Stephan, K.E., and Buhmann, J.M. (2010). The Balanced Accuracy and Its Posterior Distribution. In 2010 20th International Conference on Pattern Recognition, pp. 3121–3124.

References

Brouwer, G.J., and Heeger, D.J. (2013). Categorical clustering of the neural representation of color. J. Neurosci. 33, 15454–15465.

Brown, T.I., Carr, V.A., LaRocque, K.F., Favila, S.E., Gordon, A.M., Bowles, B., Bailenson, J.N., and Wagner, A.D. (2016). Prospective representation of navigational goals in the human hippocampus. Science 352, 1323–1326.

Brun, V.H., Solstad, T., Kjelstrup, K.B., Fyhn, M., Witter, M.P., Moser, E.I., and Moser, M.-B. (2008). Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. Hippocampus 18, 1200–1212.

Buckner, R.L., Krienen, F.M., and Yeo, B.T.T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. Nat. Neurosci. 16, 832–837.

Burgess, N., Maguire, E.A., and O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. Neuron 35, 625–641.

Bush, D., Bisby, J.A., Bird, C.M., Gollwitzer, S., Rodionov, R., Diehl, B., McEvoy, A.W., Walker, M.C., and Burgess, N. (2017). Human hippocampal theta power indicates movement onset and distance travelled. Proc. Natl. Acad. Sci. U. S. A. 114, 12297–12302.

Chadwick, M.J., Jolly, A.E.J., Amos, D.P., Hassabis, D., and Spiers, H.J. (2015). A Goal Direction Signal in the Human Entorhinal/Subicular Region. Current Biology 25, 87–92.

Chen, G., King, J.A., Burgess, N., and O'Keefe, J. (2013). How vision and movement combine in the hippocampal place code. Proc. Natl. Acad. Sci. U. S. A. 110, 378–383.

Chen, L.L., Lin, L.H., Green, E.J., Barnes, C.A., and McNaughton, B.L. (1994a). Head-direction cells in the rat posterior cortex. I. Anatomical distribution and behavioral modulation. Exp. Brain Res. 101, 8–23.

Chen, L.L., Lin, L.H., Barnes, C.A., and McNaughton, B.L. (1994b). Head-direction cells in the rat posterior cortex. II. Contributions of visual and ideothetic information to the directional firing. Exp. Brain Res. 101, 24–34.

Chrastil, E.R., and Warren, W.H. (2014). From cognitive maps to cognitive graphs. PLoS One 9, e112544.

Chrastil, E.R., Sherrill, K.R., Hasselmo, M.E., and Stern, C.E. (2015). There and Back Again: Hippocampus and Retrosplenial Cortex Track Homing Distance during Human Path Integration. J. Neurosci. 35, 15442–15452.

Chrastil, E.R., Sherrill, K.R., Hasselmo, M.E., and Stern, C.E. (2016). Which way and how far? Tracking of translation and rotation information for human path integration. Hum. Brain Mapp. 37, 3636–3655.

Ciaramelli, E. (2008). The role of ventromedial prefrontal cortex in navigation: A case of impaired wayfinding and rehabilitation. Neuropsychologia 46, 2099–2105.

Connor, C.E., Preddie, D.C., Gallant, J.L., and Van Essen, D.C. (1997). Spatial attention effects in macaque area V4. J. Neurosci. 17, 3201–3214.

References

Corbetta, M., Akbudak, E., Conturo, T.E., Snyder, A.Z., Ollinger, J.M., Drury, H.A., Linenweber, M.R., Petersen, S.E., Raichle, M.E., Van Essen, D.C., et al. (1998). A common network of functional areas for attention and eye movements. Neuron 21, 761–773.

Cox, D.D., and Savoy, R.L. (2003). Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. NeuroImage 19, 261–270.

Cremers, H.R., Wager, T.D., and Yarkoni, T. (2017). The relation between statistical power and inference in fMRI. PLoS One 12, e0184923.

Çukur, T., Nishimoto, S., Huth, A.G., and Gallant, J.L. (2013). Attention during natural vision warps semantic representation across the human brain. Nat. Neurosci. 16, 763–770.

Çukur, T., Huth, A.G., Nishimoto, S., and Gallant, J.L. (2016). Functional Subdomains within Scene-Selective Cortex: Parahippocampal Place Area, Retrosplenial Complex, and Occipital Place Area. The Journal of Neuroscience 36, 10257–10273.

Dabaghian, Y., Brandt, V.L., and Frank, L.M. (2014). Reconceiving the hippocampal map as a topological template. Elife 3, e03476.

Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9, 179–194.

Danjo, T., Toyoizumi, T., and Fujisawa, S. (2018). Spatial representations of self and other in the hippocampus. Science 359, 213–218.

David, S.V., Hayden, B.Y., Mazer, J.A., and Gallant, J.L. (2008). Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. Neuron 59, 509–521.

Deniz, F., Nunez-Elizalde, A.O., Huth, A.G., and Gallant, J.L. (2019). The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. J. Neurosci. 39, 7722–7736.

Deshmukh, S.S., and Knierim, J.J. (2011). Representation of non-spatial and spatial information in the lateral entorhinal cortex. Front. Behav. Neurosci. 5, 69.

Deshmukh, S.S., and Knierim, J.J. (2013). Influence of local objects on hippocampal representations: Landmark vectors and memory. Hippocampus 23, 253–267.

Deshmukh, S.S., Johnson, J.L., and Knierim, J.J. (2012). Perirhinal cortex represents nonspatial, but not spatial, information in rats foraging in the presence of objects: comparison with lateral entorhinal cortex. Hippocampus 22, 2045–2058.

Desmond, J.E., and Glover, G.H. (2002). Estimating sample size in functional MRI (fMRI) neuroimaging studies: statistical power analyses. J. Neurosci. Methods 118, 115–128.

Dilks, D.D., Julian, J.B., Paunov, A.M., and Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. J. Neurosci. 33, 1331–6a.

Doeller, C.F., Barry, C., and Burgess, N. (2010). Evidence for grid cells in a human memory network. Nature 463, 657–661.

References

Dosovitskiy, A., Ros, G., Codevilla, F., Lopez, A., and Koltun, V. (2017). CARLA: An open urban driving simulator.

Downing, P.E., Jiang, Y., Shuman, M., and Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. Science 293, 2470–2473.

Dupre la Tour T Eickenberg ME Gallant, J.L. Variance decomposition with banded ridge regression. In Prep.

Ekstrom, A.D., Kahana, M.J., Caplan, J.B., Fields, T.A., Isham, E.A., Newman, E.L., and Fried, I. (2003). Cellular networks underlying human spatial navigation. Nature 425, 184–188.

Ekstrom, A.D., Arnold, A.E.G.F., and Iaria, G. (2014). A critical review of the allocentric spatial representation and its neural underpinnings: toward a network-based perspective. Front. Hum. Neurosci. 8, 803.

Epstein, R.A. (2008). Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn. Sci. 12, 388–396.

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

Epstein, R., Harris, A., Stanley, D., and Kanwisher, N. (1999). The parahippocampal place area: recognition, navigation, or encoding? Neuron 23, 115–125.

Epstein, R., Graham, K.S., and Downing, P.E. (2003). Viewpoint-specific scene representations in human parahippocampal cortex. Neuron 37, 865–876.

Epstein, R.A., Patai, E.Z., Julian, J.B., and Spiers, H.J. (2017). The cognitive map in humans: spatial navigation and beyond. Nat. Neurosci. 20, 1504–1513.

Feierstein, C.E., Quirk, M.C., Uchida, N., Sosulski, D.L., and Mainen, Z.F. (2006). Representation of spatial goals in rat orbitofrontal cortex. Neuron 51, 495–507.

Field, D.J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. J. Opt. Soc. Am. A 4, 2379–2394.

Fischer, L.F., Mojica Soto-Albors, R., Buck, F., and Harnett, M.T. (2020). Representation of visual landmarks in retrosplenial cortex. Elife 9.

Fitzpatrick, D.C., Batra, R., Stanford, T.R., and Kuwada, S. (1997). A neuronal population code for sound localization. Nature 388, 871–874.

Friston, K.J. (1994a). Statistical parametric mapping. In Functional Neuroimaging: Technical Foundations , (Pp, R.W. Thatcher, ed. (San Diego, CA, US: Academic Press, xx), pp. 79–93.

Friston, K.J. (1994b). Functional and effective connectivity in neuroimaging: A synthesis. Hum. Brain Mapp. 2, 56–78.

Fyhn, M., Molden, S., Witter, M.P., Moser, E.I., and Moser, M.-B. (2004). Spatial representation in the entorhinal cortex. Science 305, 1258–1264.

Gao, J.S., Huth, A.G., Lescroart, M.D., and Gallant, J.L. (2015). Pycortex: an interactive surface visualizer for fMRI. Front. Neuroinform. 9, 23.

References

Garling, T., Book, A., and Lindberg, E. (1984). Cognitive mapping of large-scale environments. Environ. Behav. 16, 3–34.

Gauthier, I., Skudlarski, P., Gore, J.C., and Anderson, A.W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. Nat. Neurosci. 3, 191–197.

Gazova, I., Vlcek, K., Laczó, J., Nedelska, Z., Hyncicova, E., Mokrisova, I., Sheardova, K., and Hort, J. (2012). Spatial navigation-a unique window into physiological and pathological aging. Front. Aging Neurosci. 4, 16.

Ghaem, O., Mellet, E., Crivello, F., Tzourio, N., Mazoyer, B., Berthoz, A., and Denis, M. (1997). Mental navigation along memorized routes activates the hippocampus, precuneus, and insula. Neuroreport 8, 739–744.

Glover, G.H. (1999). Deconvolution of Impulse Response in Event-Related BOLD fMRI1. Neuroimage 9, 416–429.

Gordon, E.M., Breeden, A.L., Bean, S.E., and Vaidya, C.J. (2014). Working memory-related changes in functional connectivity persist beyond task disengagement. Hum. Brain Mapp. 35, 1004–1017.

Gourtzelidis, P., Tzagarakis, C., Lewis, S.M., Crowe, D.A., Auerbach, E., Jerde, T.A., Uğurbil, K., and Georgopoulos, A.P. (2005). Mental maze solving: directional fMRI tuning and population coding in the superior parietal lobule. Exp. Brain Res. 165, 273–282.

Granon, S., and Poucet, B. (1995). Medial prefrontal lesions in the rat and spatial navigation: Evidence for impaired planning. Behavioral Neuroscience 109, 474–484.

Grill-Spector, K. (2003). The neural basis of object perception. Curr. Opin. Neurobiol. 13, 399.

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.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E.I. (2005). Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801–806.

Hansen, K.A., Kay, K.N., and Gallant, J.L. (2007). Topographic organization in and near human visual area V4. J. Neurosci. 27, 11896–11911.

Harel, A., Kravitz, D.J., and Baker, C.I. (2014). Task context impacts visual object processing differentially across the cortex. Proc. Natl. Acad. Sci. U. S. A. 111, E962-71.

Harvey, C.D., Collman, F., Dombeck, D.A., and Tank, D.W. (2009). Intracellular dynamics of hippocampal place cells during virtual navigation. Nature 461, 941–946.

Harvey, C.D., Coen, P., and Tank, D.W. (2012). Choice-specific sequences in parietal cortex during a virtual-navigation decision task. Nature 484, 62–68.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–1640.

References

Hasson, U., Malach, R., and Heeger, D.J. (2010). Reliability of cortical activity during natural stimulation. Trends Cogn. Sci. 14, 40–48.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293, 2425–2430.

Heeger, D.J., and Ress, D. (2002). What does fMRI tell us about neuronal activity? Nat. Rev. Neurosci. 3, 142–151.

Hok, V., Save, E., Lenck-Santini, P.P., and Poucet, B. (2005). Coding for spatial goals in the prelimbic/infralimbic area of the rat frontal cortex. Proc. Natl. Acad. Sci. U. S. A. 102, 4602–4607.

Hölscher, C., Schnee, A., Dahmen, H., Setia, L., and Mallot, H.A. (2005). Rats are able to navigate in virtual environments. J. Exp. Biol. 208, 561–569.

Howard, L.R., Javadi, A.H., Yu, Y., Mill, R.D., Morrison, L.C., Knight, R., Loftus, M.M., Staskute, L., and Spiers, H.J. (2014). The hippocampus and entorhinal cortex encode the path and Euclidean distances to goals during navigation. Curr. Biol. 24, 1331–1340.

Høydal, Ø.A., Skytøen, E.R., Andersson, S.O., Moser, M.-B., and Moser, E.I. (2019). Object-vector coding in the medial entorhinal cortex. Nature 568, 400–404.

Huth, A.G., Nishimoto, S., Vu, A.T., and Gallant, J.L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron 76, 1210–1224.

Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., and Gallant, J.L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532, 453–458.

Iaria, G., Chen, J.-K., Guariglia, C., Ptito, A., and Petrides, M. (2007). Retrosplenial and hippocampal brain regions in human navigation: complementary functional contributions to the formation and use of cognitive maps. Eur. J. Neurosci. 25, 890–899.

Jacob, P.-Y., Casali, G., Spieser, L., Page, H., Overington, D., and Jeffery, K. (2017). An independent, landmark-dominated head-direction signal in dysgranular retrosplenial cortex. Nat. Neurosci. 20, 173–175.

Jacobs, J., Weidemann, C.T., Miller, J.F., Solway, A., Burke, J.F., Wei, X.-X., Suthana, N., Sperling, M.R., Sharan, A.D., Fried, I., et al. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. Nat. Neurosci. 16, 1188–1190.

Jaderberg, M., Czarnecki, W.M., Dunning, I., Marris, L., Lever, G., Castañeda, A.G., Beattie, C., Rabinowitz, N.C., Morcos, A.S., Ruderman, A., et al. (2019). Human-level performance in 3D multiplayer games with population-based reinforcement learning. Science 364, 859–865.

Javadi, A.-H., Emo, B., Howard, L.R., Zisch, F.E., Yu, Y., Knight, R., Pinelo Silva, J., and Spiers, H.J. (2017). Hippocampal and prefrontal processing of network topology to simulate the future. Nat. Commun. 8, 14652.

References

Jenkinson, M., and Smith, S. (2001). The role of registration in functional magnetic resonance imaging. In Medical Image Registration, (CRC Press), pp. 183–198.

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825–841.

Julian, J.B., Keinath, A.T., Muzzio, I.A., and Epstein, R.A. (2015). Place recognition and heading retrieval are mediated by dissociable cognitive systems in mice. Proc. Natl. Acad. Sci. U. S. A. 112, 6503–6508.

Julian, J.B., Ryan, J., Hamilton, R.H., and Epstein, R.A. (2016). The Occipital Place Area Is Causally Involved in Representing Environmental Boundaries during Navigation. Curr. Biol. 26, 1104–1109.

Kamitani, Y., and Tong, F. (2005). Decoding the visual and subjective contents of the human brain. Nat. Neurosci. 8, 679–685.

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

Kanwisher, N., McDermott, J., and Chun, M.M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. J. Neurosci. 17, 4302–4311.

Kaplan, R., King, J., Koster, R., Penny, W.D., Burgess, N., and Friston, K.J. (2017). The Neural Representation of Prospective Choice during Spatial Planning and Decisions. PLoS Biol. 15, e1002588.

Kastner, S., Pinsk, M.A., De Weerd, P., Desimone, R., and Ungerleider, L.G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. Neuron 22, 751–761.

Kay, K.N., David, S.V., Prenger, R.J., Hansen, K.A., and Gallant, J.L. (2008a). Modeling low-frequency fluctuation and hemodynamic response timecourse in event-related fMRI. Human Brain Mapping 29, 142–156.

Kay, K.N., Naselaris, T., Prenger, R.J., and Gallant, J.L. (2008b). Identifying natural images from human brain activity. Nature 452, 352–355.

Kim, M., and Maguire, E.A. (2018). Encoding of 3D head direction information in the human brain. Hippocampus.

Kjelstrup, K.B., Solstad, T., Brun, V.H., Hafting, T., Leutgeb, S., Witter, M.P., Moser, E.I., and Moser, M.-B. (2008). Finite scale of spatial representation in the hippocampus. Science 321, 140–143.

Kobayashi, Y., and Amaral, D.G. (2003). Macaque monkey retrosplenial cortex: II. Cortical afferents. J. Comp. Neurol. 466, 48–79.

Kobayashi, Y., and Amaral, D.G. (2007). Macaque monkey retrosplenial cortex: III. Cortical efferents. J. Comp. Neurol. 502, 810–833.

References

Krakauer, J.W., Ghazanfar, A.A., Gomez-Marin, A., MacIver, M.A., and Poeppel, D. (2017). Neuroscience Needs Behavior: Correcting a Reductionist Bias. Neuron 93, 480–490.

Kriegeskorte, N., Mur, M., and Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. Front. Syst. Neurosci. 2, 4.

Kullback, S., and Leibler, R.A. (1951). On Information and Sufficiency. Ann. Math. Stat. 22, 79–86.

Laczó, J., Parizkova, M., and Moffat, S.D. (2018). Spatial navigation, aging and Alzheimer's disease. Aging (Albany NY) 10, 3050–3051.

Ledergerber, D., Battistin, C., Blackstad, J.S., Gardner, R.J., Witter, M.P., Moser, M.-B., Roudi, Y., and Moser, E.I. (2021). Task-dependent mixed selectivity in the subiculum. Cell Rep. 35, 109175.

Lescroart, M.D., and Gallant, J.L. (2019). Human Scene-Selective Areas Represent 3D Configurations of Surfaces. Neuron 101, 178-192.e7.

Lescroart, M.D., Stansbury, D.E., and Gallant, J.L. (2015). Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. Front. Comput. Neurosci. 9, 135.

Lester, A.W., Moffat, S.D., Wiener, J.M., Barnes, C.A., and Wolbers, T. (2017). The aging navigational system. Neuron 95, 1019–1035.

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., and Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. J. Neurosci. 29, 9771–9777.

Lin, J. (1991). Divergence measures based on the Shannon entropy. IEEE Trans. Inf. Theory 37, 145–151.

Luck, S.J., Chelazzi, L., Hillyard, S.A., and Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. J. Neurophysiol. 77, 24–42.

Maguire, E.A. (2001). The retrosplenial contribution to human navigation: a review of lesion and neuroimaging findings. Scand. J. Psychol. 42, 225–238.

Maguire, E.A., Woollett, K., and Spiers, H.J. (2006). London taxi drivers and bus drivers: a structural MRI and neuropsychological analysis. Hippocampus 16, 1091–1101.

Mante, V., Sussillo, D., Shenoy, K.V., and Newsome, W.T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503, 78–84.

Mao, D., Avila, E., Caziot, B., Laurens, J., Dickman, J.D., and Angelaki, D.E. (2020). Spatial representations in macaque hippocampal formation (bioRxiv).

Marchette, S.A., Vass, L.K., Ryan, J., and Epstein, R.A. (2014). Anchoring the neural compass: coding of local spatial reference frames in human medial parietal lobe. Nat. Neurosci. 17, 1598–1606.

References

Mathiak, K., and Weber, R. (2006). Toward brain correlates of natural behavior: fMRI during violent video games. Human Brain Mapping 27, 948–956.

Mathiak, K.A., Klasen, M., Weber, R., Ackermann, H., Shergill, S.S., and Mathiak, K. (2011). Reward system and temporal pole contributions to affective evaluation during a first person shooter video game. BMC Neurosci. 12, 66.

Matusz, P.J., Dikker, S., Huth, A.G., and Perrodin, C. (2019). Are We Ready for Real-world Neuroscience? J. Cogn. Neurosci. 31, 327–338.

Mazer, J.A., and Gallant, J.L. (2003). Goal-related activity in V4 during free viewing visual search. Neuron 40, 1241–1250.

McAdams, C.J., and Maunsell, J.H.R. (1999). Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. J. Neurosci. 19, 431–441.

Mcnaughton, B.L., Leonard, B., and Chen, L. (1989). Cortical-hippocampal interactions and cognitive mapping: A hypothesis based on reintegration of the parietal and inferotemporal pathways for visual processing. Psychobiology 17, 230–235.

Miller, G.A. (1995). WordNet. Commun. ACM 38, 39–41.

Mizumori, S.J., and Williams, J.D. (1993). Directionally selective mnemonic properties of neurons in the lateral dorsal nucleus of the thalamus of rats. J. Neurosci. 13, 4015–4028.

Montello, D.R. (1998). A new framework for understanding the acquisition of spatial knowledge in large-scale environments. pp. 143–154.

Moore, T., and Fallah, M. (2001). Control of eye movements and spatial attention. Proc. Natl. Acad. Sci. U. S. A. 98, 1273–1276.

Muller, R.U., and Kubie, J.L. (1987). The effects of changes in the environment on the spatial firing of hippocampal complex-spike cells. J. Neurosci. 7, 1951–1968.

Mumford, J.A., and Nichols, T.E. (2008). Power calculation for group fMRI studies accounting for arbitrary design and temporal autocorrelation. Neuroimage 39, 261–268.

Naselaris, T., Prenger, R.J., Kay, K.N., Oliver, M., and Gallant, J.L. (2009). Bayesian reconstruction of natural images from human brain activity. Neuron 63, 902–915.

Naselaris, T., Kay, K.N., Nishimoto, S., and Gallant, J.L. (2011). Encoding and decoding in fMRI. Neuroimage 56, 400–410.

Nastase, S.A., Connolly, A.C., Oosterhof, N.N., Halchenko, Y.O., Guntupalli, J.S., Visconti di Oleggio Castello, M., Gors, J., Gobbini, M.I., and Haxby, J.V. (2017). Attention Selectively Reshapes the Geometry of Distributed Semantic Representation. Cereb. Cortex 27, 4277–4291.

Nau, M., Navarro Schröder, T., Bellmund, J.L.S., and Doeller, C.F. (2018). Hexadirectional coding of visual space in human entorhinal cortex. Nat. Neurosci. 21, 188–190.

References

Nishimoto, S., and Gallant, J.L. (2011). A three-dimensional spatiotemporal receptive field model explains responses of area MT neurons to naturalistic movies. J. Neurosci. 31, 14551–14564.

Nishimoto, S., Vu, A.T., Naselaris, T., Benjamini, Y., Yu, B., and Gallant, J.L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. Curr. Biol. 21, 1641–1646.

Nitz, D. (2009). Parietal cortex, navigation, and the construction of arbitrary reference frames for spatial information. Neurobiol. Learn. Mem. 91, 179–185.

Nitz, D.A. (2006). Tracking route progression in the posterior parietal cortex. Neuron 49, 747–756.

Nitz, D.A. (2012). Spaces within spaces: rat parietal cortex neurons register position across three reference frames. Nat. Neurosci. 15, 1365–1367.

Norman, K.A., Polyn, S.M., Detre, G.J., and Haxby, J.V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn. Sci. 10, 424–430.

Nunez-Elizalde, A.O., Huth, A.G., and Gallant, J.L. (2019). Voxelwise encoding models with non-spherical multivariate normal priors. Neuroimage 197, 482–492.

Ogawa, S., Lee, T.M., Kay, A.R., and Tank, D.W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc. Natl. Acad. Sci. U. S. A. 87, 9868–9872.

O'Keefe, J., and Conway, D.H. (1978). Hippocampal place units in the freely moving rat: why they fire where they fire. Exp. Brain Res. 31, 573–590.

O'Keefe, J., and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. Brain Res. 34, 171–175.

O'Keefe, J., and Nadal, L. (1978). The Hippocampus as a Cognitive Map: Oxford University.

O'Keefe, J., and Speakman, A. (1987). Single unit activity in the rat hippocampus during a spatial memory task. Exp. Brain Res. 68, 1–27.

Omer, D.B., Maimon, S.R., Las, L., and Ulanovsky, N. (2018). Social place-cells in the bat hippocampus. Science 359, 218–224.

Park, S., and Chun, M.M. (2009). Different roles of the parahippocampal place area (PPA) and retrosplenial cortex (RSC) in panoramic scene perception. Neuroimage 47, 1747–1756.

Parslow, D.M., Rose, D., Brooks, B., Fleminger, S., Gray, J.A., Giampietro, V., Brammer, M.J., Williams, S., Gasston, D., Andrew, C., et al. (2004). Allocentric spatial memory activation of the hippocampal formation measured with fMRI. Neuropsychology 18, 450–461.

Paus, T. (1996). Location and function of the human frontal eye-field: a selective review. Neuropsychologia 34, 475–483.

Peirce, J.W. (2007). PsychoPy--psychophysics software in Python. J. Neurosci. Methods 162, 8–13.

References

Persichetti, A.S., and Dilks, D.D. (2018). Dissociable Neural Systems for Recognizing Places and Navigating through Them. J. Neurosci. 38, 10295–10304.

Pillow, J.W., Shlens, J., Paninski, L., Sher, A., Litke, A.M., Chichilnisky, E.J., and Simoncelli, E.P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. Nature 454, 995–999.

Pinto, L., Rajan, K., DePasquale, B., Thiberge, S.Y., Tank, D.W., and Brody, C.D. (2019). Task-dependent changes in the large-scale dynamics and necessity of cortical regions. Neuron 104, 810-824.e9.

Pouget, A., Dayan, P., and Zemel, R. (2000). Information processing with population codes. Nature Reviews Neuroscience 1, 125–132.

Power, J.D., Silver, B.M., Silverman, M.R., Ajodan, E.L., Bos, D.J., and Jones, R.M. (2019). Customized head molds reduce motion during resting state fMRI scans. Neuroimage 189, 141–149.

Reich, D.S., Mechler, F., and Victor, J.D. (2001). Independent and redundant information in nearby cortical neurons. Science 294, 2566–2568.

Reynolds, J.H., Pasternak, T., and Desimone, R. (2000). Attention increases sensitivity of V4 neurons. Neuron 26, 703–714.

Rosenbaum, R.S., Ziegler, M., Winocur, G., Grady, C.L., and Moscovitch, M. (2004). "I have often walked down this street before": fMRI studies on the hippocampus and other structures during mental navigation of an old environment. Hippocampus 14, 826–835.

Rossi, M.A., Hayrapetyan, V.Y., Maimon, B., Mak, K., Je, H.S., and Yin, H.H. (2012). Prefrontal cortical mechanisms underlying delayed alternation in mice. J. Neurophysiol. 108, 1211–1222.

Sarel, A., Finkelstein, A., Las, L., and Ulanovsky, N. (2017). Vectorial representation of spatial goals in the hippocampus of bats. Science 355, 176–180.

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B.L., Witter, M.P., Moser, M.-B., and Moser, E.I. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. Science 312, 758–762.

Sato, N., Sakata, H., Tanaka, Y.L., and Taira, M. (2006). Navigation-associated medial parietal neurons in monkeys. Proc. Natl. Acad. Sci. U. S. A. 103, 17001–17006.

Savelli, F., Yoganarasimha, D., and Knierim, J.J. (2008). Influence of boundary removal on the spatial representations of the medial entorhinal cortex. Hippocampus 18, 1270–1282.

Savelli, F., Luck, J.D., and Knierim, J.J. (2017). Framing of grid cells within and beyond navigation boundaries. Elife 6.

Saxe, R., and Kanwisher, N. (2003). People thinking about thinking people: The role of the temporo-parietal junction in "theory of mind." Neuroimage 19, 1835–1842.

References

Shen, H., Wang, L., Liu, Y., and Hu, D. (2010). Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI. Neuroimage 49, 3110–3121.

Sherrill, K.R., Erdem, U.M., Ross, R.S., Brown, T.I., Hasselmo, M.E., and Stern, C.E. (2013). Hippocampus and retrosplenial cortex combine path integration signals for successful navigation. J. Neurosci. 33, 19304–19313.

Shine, J.M., Hearne, L.J., Breakspear, M., Hwang, K., Müller, E.J., Sporns, O., Poldrack, R.A., Mattingley, J.B., and Cocchi, L. (2019a). The Low-Dimensional Neural Architecture of Cognitive Complexity Is Related to Activity in Medial Thalamic Nuclei. Neuron 104, 849-855.e3.

Shine, J.M., Breakspear, M., Bell, P.T., Ehgoetz Martens, K.A., Shine, R., Koyejo, O., Sporns, O., and Poldrack, R.A. (2019b). Human cognition involves the dynamic integration of neural activity and neuromodulatory systems. Nat. Neurosci. 22, 289–296.

Shine, J.P., Valdés-Herrera, J.P., Hegarty, M., and Wolbers, T. (2016). The Human Retrosplenial Cortex and Thalamus Code Head Direction in a Global Reference Frame. J. Neurosci. 36, 6371–6381.

Silver, M.A., Ress, D., and Heeger, D.J. (2005). Topographic Maps of Visual Spatial Attention in Human Parietal Cortex. Journal of Neurophysiology 94, 1358–1371.

Smith, S.M., Vidaurre, D., Beckmann, C.F., Glasser, M.F., Jenkinson, M., Miller, K.L., Nichols, T.E., Robinson, E.C., Salimi-Khorshidi, G., Woolrich, M.W., et al. (2013). Functional connectomics from resting-state fMRI. Trends Cogn. Sci. 17, 666–682.

Snyder, L.H., Batista, A.P., and Andersen, R.A. (1997). Coding of intention in the posterior parietal cortex. Nature 386, 167–170.

Solstad, T., Moser, E.I., and Einevoll, G.T. (2006). From grid cells to place cells: a mathematical model. Hippocampus 16, 1026–1031.

Solstad, T., Boccara, C.N., Kropff, E., Moser, M.-B., and Moser, E.I. (2008). Representation of geometric borders in the entorhinal cortex. Science 322, 1865–1868.

Song, E.Y., Kim, Y.B., Kim, Y.H., and Jung, M.W. (2005). Role of active movement in place-specific firing of hippocampal neurons. Hippocampus 15, 8–17.

Spiers, H.J., and Gilbert, S.J. (2015). Solving the detour problem in navigation: a model of prefrontal and hippocampal interactions. Front. Hum. Neurosci. 9, 125.

Spiers, H.J., and Maguire, E.A. (2006). Thoughts, behaviour, and brain dynamics during navigation in the real world. Neuroimage 31, 1826–1840.

Spiers, H.J., and Maguire, E.A. (2007a). Neural substrates of driving behaviour. Neuroimage 36, 245–255.

Spiers, H.J., and Maguire, E.A. (2007b). A navigational guidance system in the human brain. Hippocampus 17, 618–626.

References

Srinivasan, L., Eden, U.T., Willsky, A.S., and Brown, E.N. (2006). A state-space analysis for reconstruction of goal-directed movements using neural signals. Neural Comput. 18, 2465–2494.

Taube, J.S. (1995). Head direction cells recorded in the anterior thalamic nuclei of freely moving rats. J. Neurosci. 15, 70–86.

Taube, J.S., Muller, R.U., and Ranck, J.B., Jr (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis. J. Neurosci. 10, 420–435.

Tootell, R.B., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R., and Belliveau, J.W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. J. Neurosci. 15, 3215–3230.

Trees, J., Snider, J., Falahpour, M., Guo, N., Lu, K., Johnson, D.C., Poizner, H., and Liu, T.T. (2014). Game controller modification for fMRI hyperscanning experiments in a cooperative virtual reality environment. MethodsX 1, 292–299.

Tsao, D.Y., and Livingstone, M.S. (2008). Mechanisms of face perception. Annu. Rev. Neurosci. 31, 411–437.

Van Dijk, K.R.A., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., and Buckner, R.L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J. Neurophysiol. 103, 297–321.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., and WU-Minn HCP Consortium (2013). The WU-Minn Human Connectome Project: an overview. Neuroimage 80, 62–79.

Vann, S.D., Aggleton, J.P., and Maguire, E.A. (2009). What does the retrosplenial cortex do? Nat. Rev. Neurosci. 10, 792.

Vass, L.K., and Epstein, R.A. (2013). Abstract representations of location and facing direction in the human brain. J. Neurosci. 33, 6133–6142.

Vedder, L.C., Miller, A.M.P., Harrison, M.B., and Smith, D.M. (2017). Retrosplenial Cortical Neurons Encode Navigational Cues, Trajectories and Reward Locations During Goal Directed Navigation. Cereb. Cortex 27, 3713–3723.

Vlcek, K. (2011). Spatial navigation impairment in healthy aging and Alzheimer's disease. In The Clinical Spectrum of Alzheimer's Disease -The Charge Toward Comprehensive Diagnostic and Therapeutic Strategies, (InTech), p.

Wang, C., Chen, X., Lee, H., Deshmukh, S.S., Yoganarasimha, D., Savelli, F., and Knierim, J.J. (2018). Egocentric coding of external items in the lateral entorhinal cortex. Science 362, 945–949.

Warren, W.H., Rothman, D.B., Schnapp, B.H., and Ericson, J.D. (2017). Wormholes in virtual space: From cognitive maps to cognitive graphs. Cognition 166, 152–163.

References

Whitlock, J.R., Sutherland, R.J., Witter, M.P., Moser, M.-B., and Moser, E.I. (2008). Navigating from hippocampus to parietal cortex. Proc. Natl. Acad. Sci. U. S. A. 105, 14755–14762.

Wiener, S.I. (1993). Spatial and behavioral correlates of striatal neurons in rats performing a self-initiated navigation task. J. Neurosci. 13, 3802–3817.

Wilber, A.A., Clark, B.J., Forster, T.C., Tatsuno, M., and McNaughton, B.L. (2014). Interaction of egocentric and world-centered reference frames in the rat posterior parietal cortex. J. Neurosci. 34, 5431–5446.

Wu, M.C.-K., David, S.V., and Gallant, J.L. (2006a). Complete functional characterization of sensory neurons by system identification. Annu. Rev. Neurosci. 29, 477–505.

Wu, W., Gao, Y., Bienenstock, E., Donoghue, J.P., and Black, M.J. (2006b). Bayesian population decoding of motor cortical activity using a Kalman filter. Neural Comput. 18, 80–118.

Wyss, J.M., and Van Groen, T. (1992). Connections between the retrosplenial cortex and the hippocampal formation in the rat: a review. Hippocampus 2, 1–11.

Yarkoni, T., and Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. Perspect. Psychol. Sci. 12, 1100–1122.

Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zöllei, L., Polimeni, J.R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. 106, 1125–1165.

Yoon, T., Okada, J., Jung, M.W., and Kim, J.J. (2008). Prefrontal cortex and hippocampus subserve different components of working memory in rats. Learn. Mem. 15, 97–105.

Zeki, S., Watson, J.D., Lueck, C.J., Friston, K.J., Kennard, C., and Frackowiak, R.S. (1991). A direct demonstration of functional specialization in human visual cortex. J. Neurosci. 11, 641–649.

8 Supplementary figures

8.1. Supplementary figures for “3 Mapping navigational representations across the brain during naturalistic driving”

Supplementary Figures

Supplementary Figure 1. Individual model performances in all subjects. Banded ridge models were fit using 33 feature spaces to data in each subject. In Figure 8 we show the aggregate performance of the overall model. The overall performance could be decomposed into performances for individual models. These show performances for the individual models corresponding to the following feature spaces: affordance, beeline distance elapsed, beeline distance remaining, controls, depth, and destination-anchored vector. Model performances are shown on the same scale as Figure 8

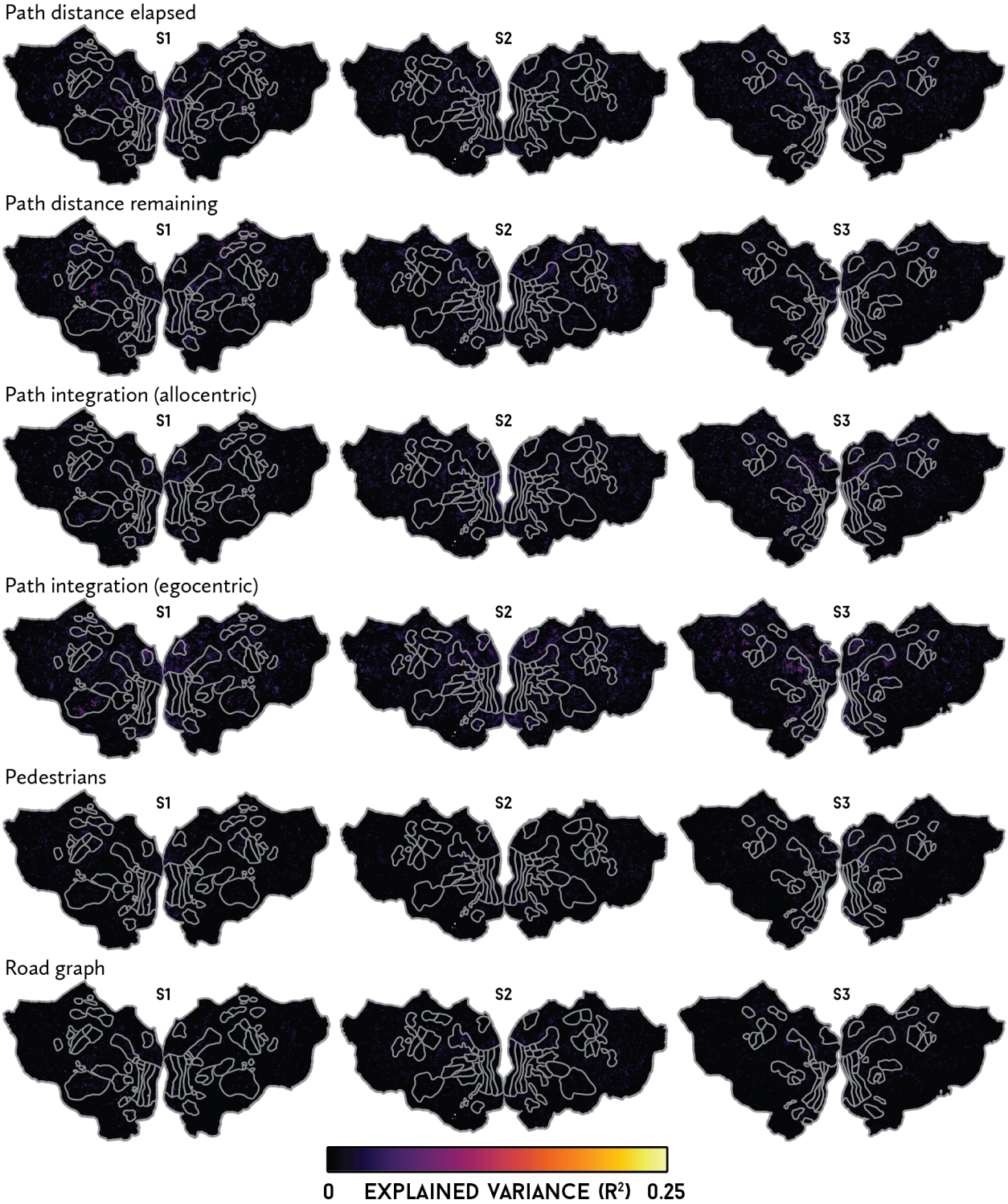
Supplementary Figures

Supplementary Figure 2. Individual model performances in all subjects. Banded ridge models were fit using 33 feature spaces to data in each subject. In Figure 8 we show the overall aggregate performance of the overall model. The overall model performance could be decomposed into performances for individual models. These show performances for the individual models corresponding to the following feature spaces: destination grid representation, destination vector, eyetracking, frame semantics, future path, and gaze direction. Model performances are shown on the same scale as Figure 8

Supplementary Figures

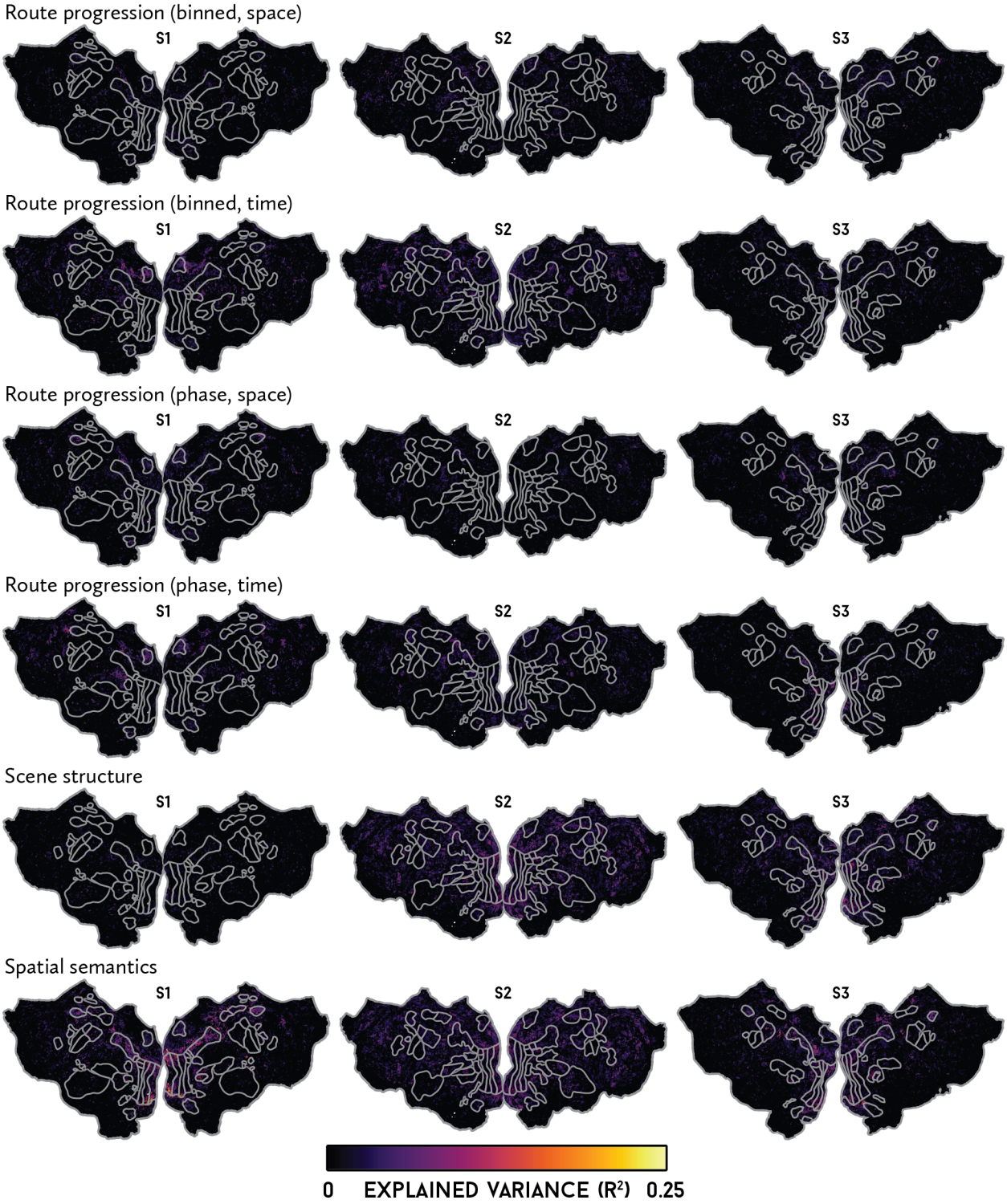
Supplementary Figure 3 Individual model performances in all subjects. Banded ridge models were fit using 33 feature spaces to data in each subject. In Figure 8 we show the overall aggregate performance of the overall model. The overall model performance could be decomposed into performances for individual models. These show performances for the individual models corresponding to the following feature spaces: gaze grid, gaze semantics, grid cells, head direction, motion-energy (frame-wise), and motion-energy (retinotopic). Model performances are shown on the same scale as Figure 8.

Supplementary Figures



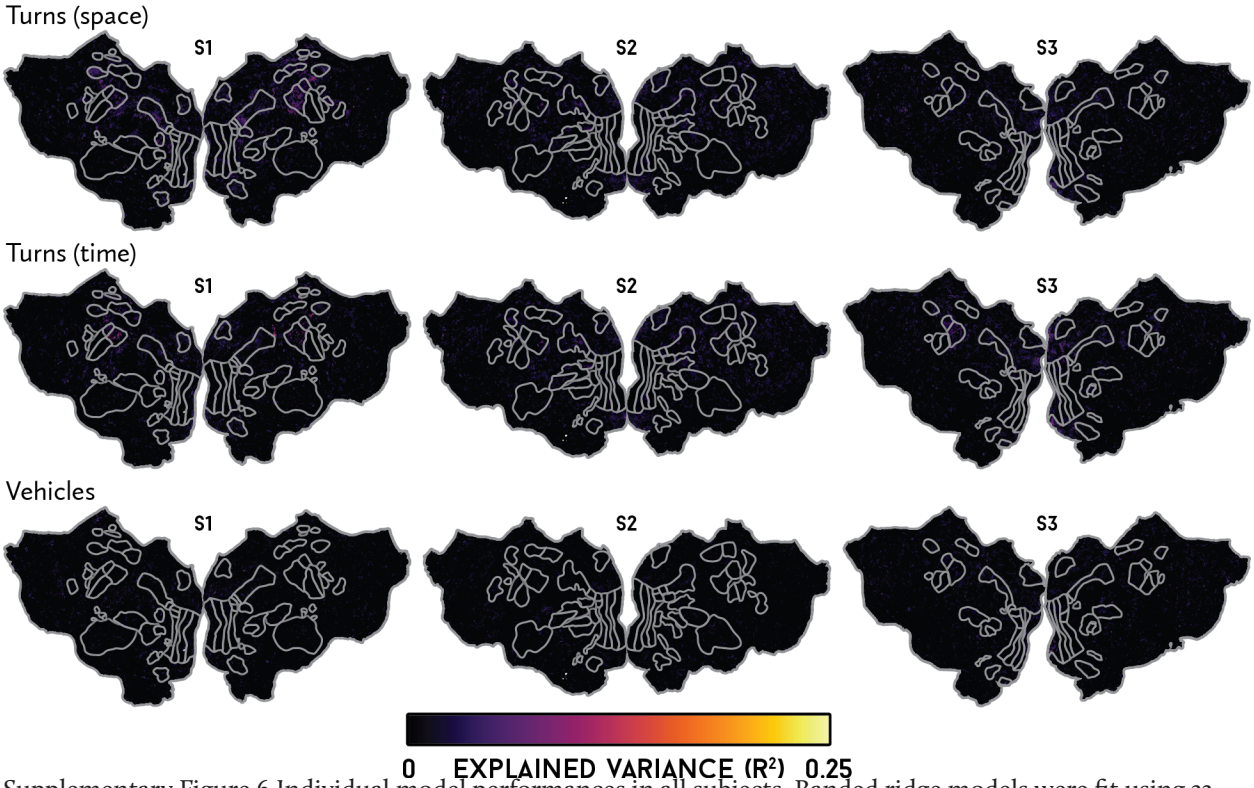
Supplementary Figure 4 Individual model performances in all subjects. Banded ridge models were fit using 33 feature spaces to data in each subject. In Figure 8 we show the overall aggregate performance of the overall model. The overall model performance could be decomposed into performances for individual models. These show performances for the individual models corresponding to the following feature spaces: path distance elapsed, path distance remaining, path integration (allocentric), path integration (egocentric), pedestrians, and road graph. Model performances are shown on the same scale as Figure 8.

Supplementary Figures



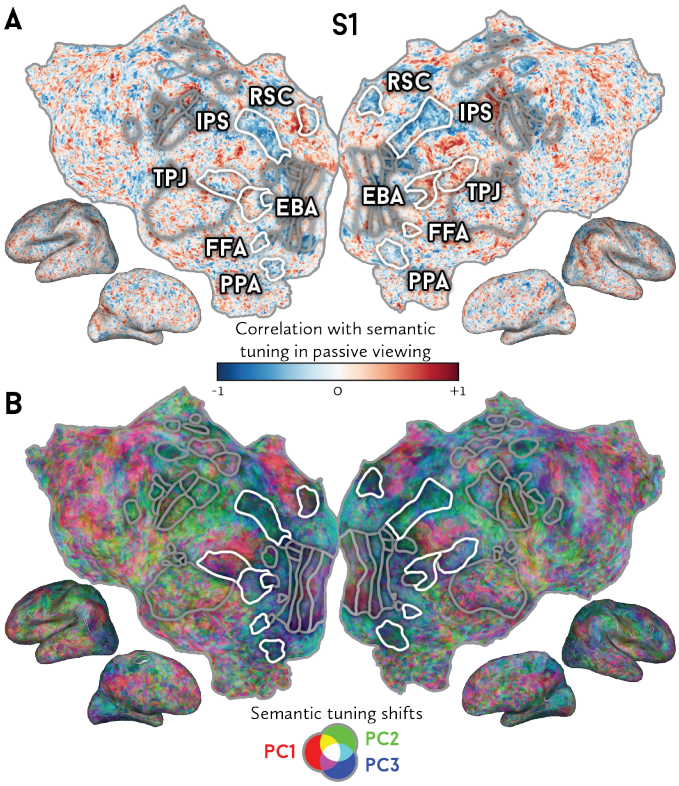
Supplementary Figure 5 Individual model performances in all subjects. Banded ridge models were fit using 33 feature spaces to data in each subject. In Figure 8 we show the overall aggregate performance of the overall model. The overall model performance could be decomposed into performances for individual models. These show performances for the individual models corresponding to the following feature spaces: binned route progression (space), binned route progression (time), route progression phase (space), route progression phase (time), scene structure, and spatial semantics. Model performances are shown on the same scale as Figure 8.

Supplementary Figures



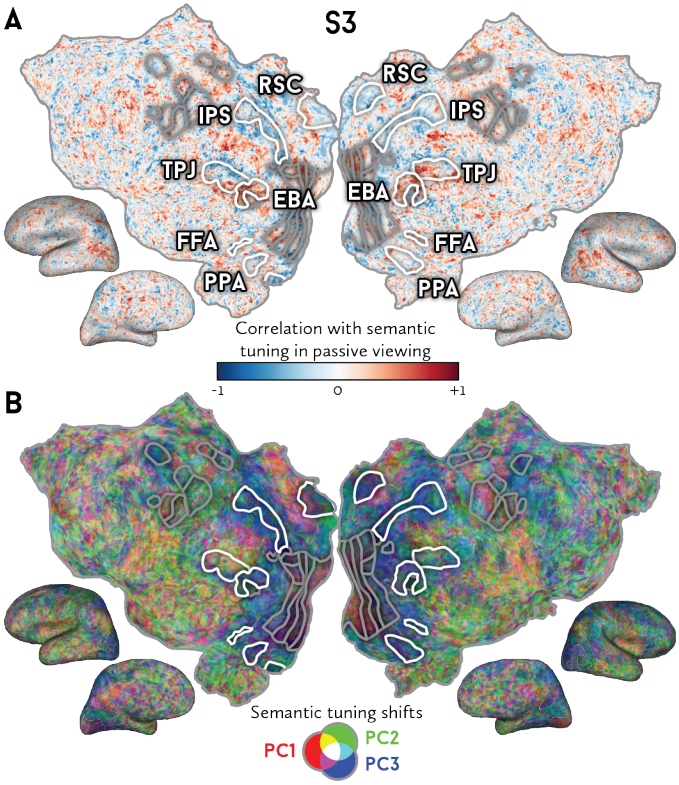
Supplementary Figure 6 Individual model performances in all subjects. Banded ridge models were fit using 33 feature spaces to data in each subject. In Figure 8 we show the overall aggregate performance of the overall model. The overall model performance could be decomposed into performances for individual models. These show performances for the individual models corresponding to the following feature spaces: turns (space), turns (time), and vehicles. Model performances are shown on the same scale as Figure 8.

8.2. Supplementary figures for "4 Active navigation elicits visual semantic tuning shifts across the cortex"



Supplementary Figure 7. Visual semantics tuning shifts for subject 1, shown in the same manner as Figure 14. A) Correlation of semantic tuning between the active driving and passive movie tasks across the cortex in subject 1. B) The first three PCs of tuning shift directions. These correspond to the same PCs shown in Figure 14C.

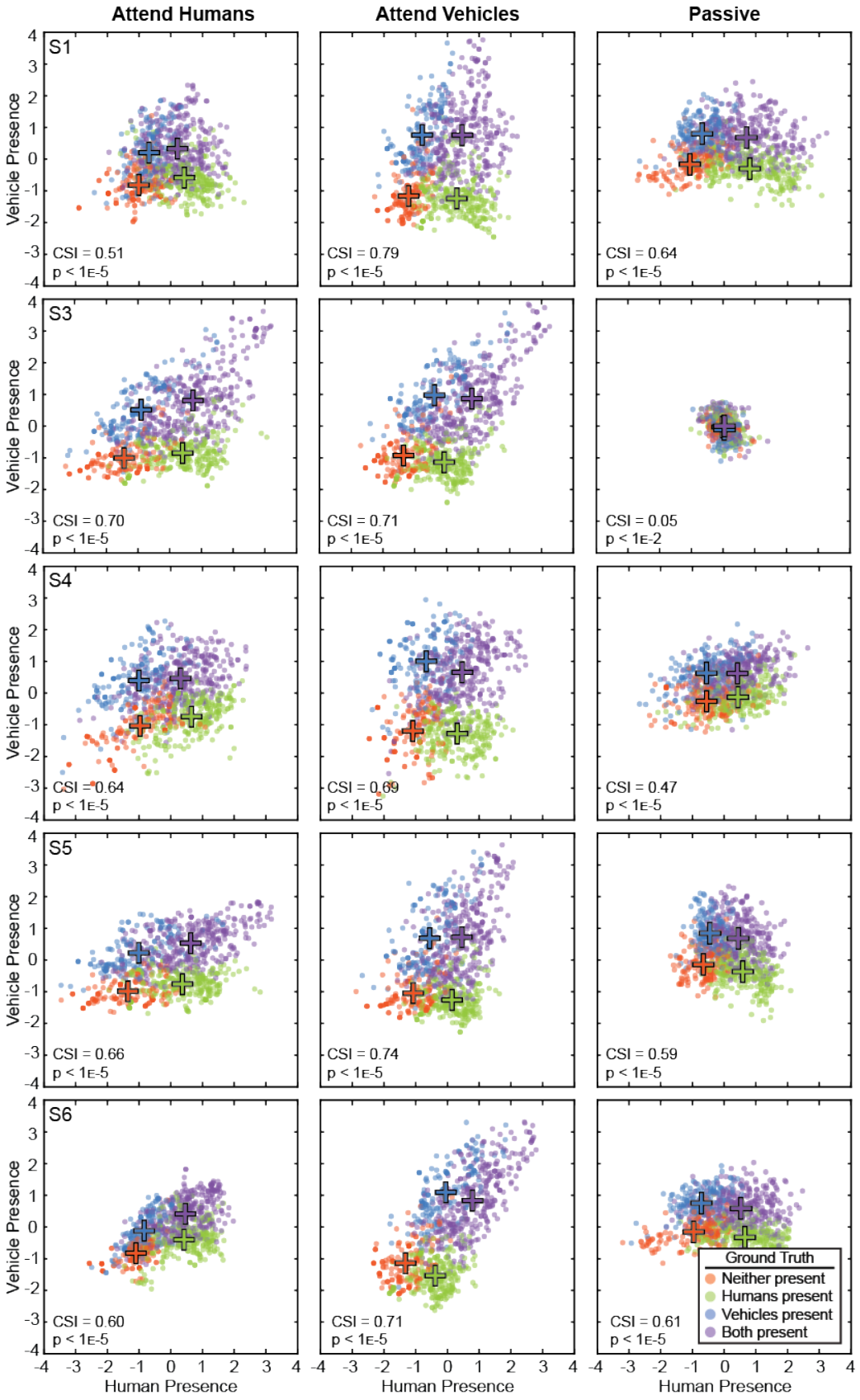
Supplementary Figures



Supplementary Figure 8. Visual semantics tuning shifts for subject 3, shown in the same manner as Figure 14. A) Correlation of semantic tuning between the active driving and passive movie tasks across the cortex in subject 3. B) The first three PCs of tuning shift directions. These correspond to the same PCs shown in Figure 14C.

8.3. Supplementary figures for "6 Voxel-based state space modeling recovers task-related cognitive states in naturalistic fMRI experiments"

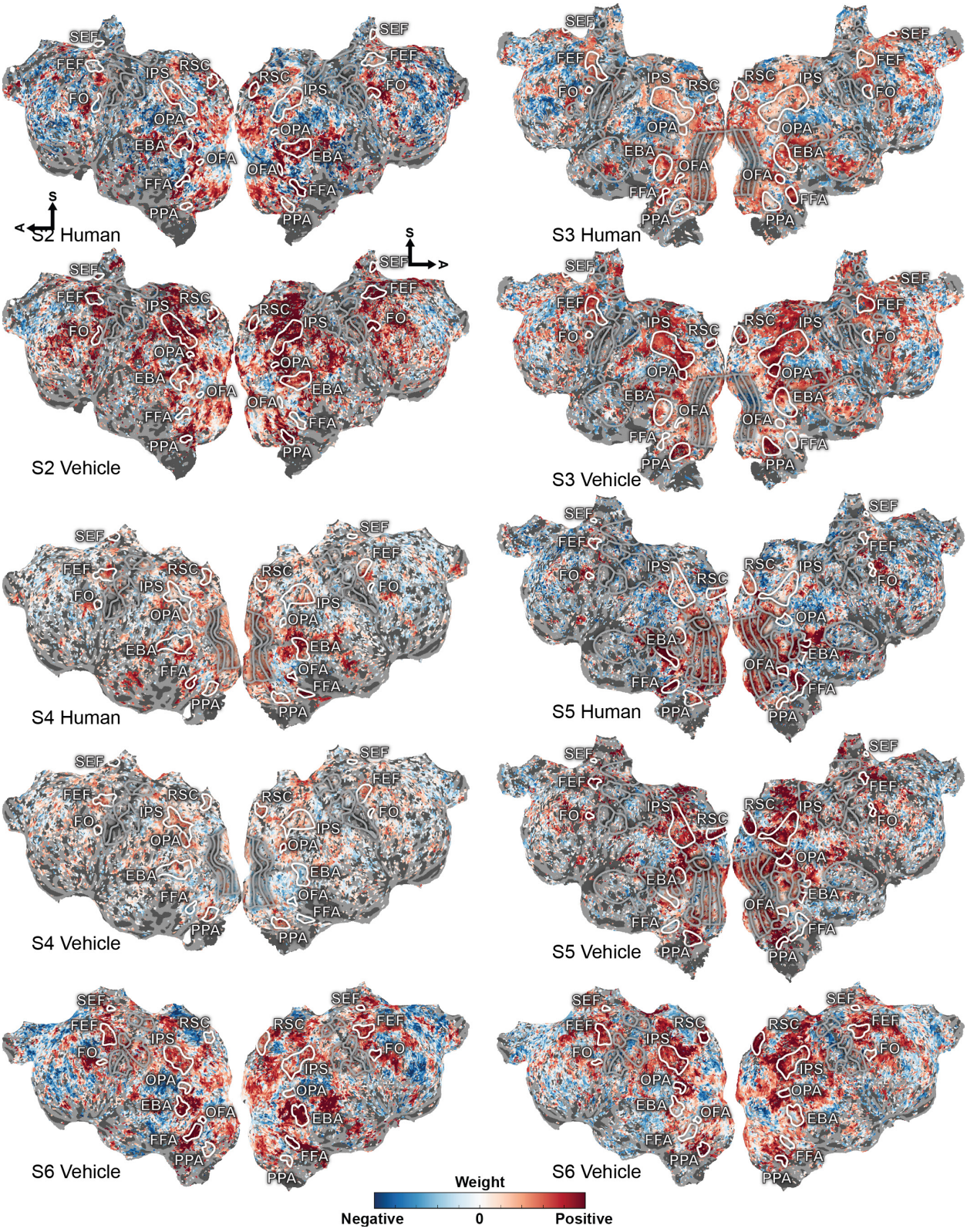
Supplementary Figures



Supplementary Figures

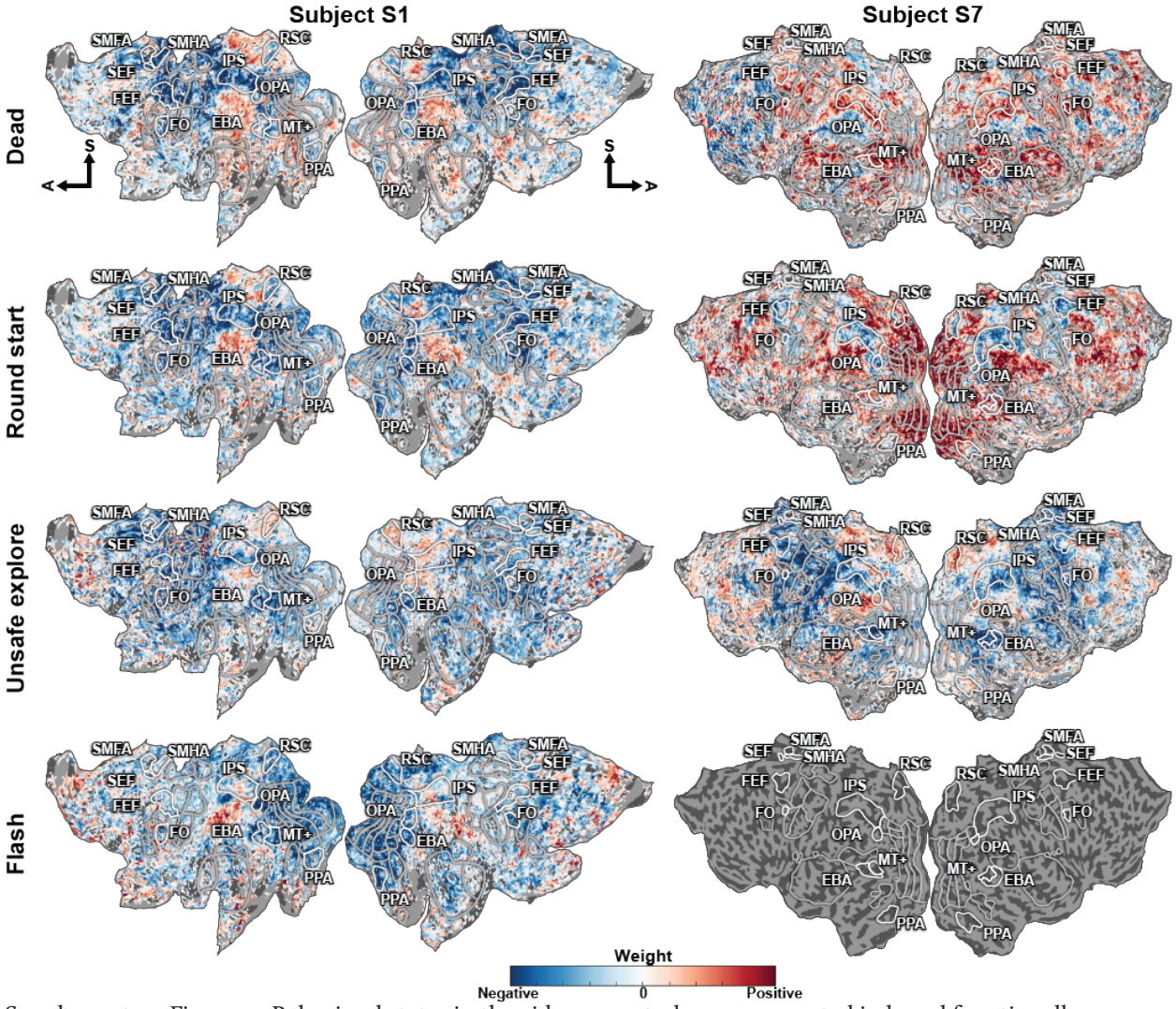
Supplementary Figure 9. (Previous page) A task-related state space for all subjects in the visual attention task. A task-related state space is recovered separated for each subject. Cortical activity at each TR for subject is projected to a point in that subject's space. The state space is projected on to the 2D plane spanned by the "human presence" (horizontal) and "vehicle presence" (vertical) axes. Each row corresponds to the state space for one subject. Projections are shown for the attend-human (left column), attend-vehicles (center column), and passive (right column) viewing conditions. Positive values indicate presence, and negative values indicate absence. TRs are color-coded by their ground truth states. Orange: neither present; blue: only humans present; green: only vehicles present; purple: both humans and vehicles present. Crosses indicate the mean positions of each group. CSIs are shows for each condition in each subject and are significant ($p < 1e\text{-}2$ or less, permutation test). Clusters are more distinct in the attentive conditions than in the passive condition ($p < 1e\text{-}5$, permutation test). These results are consistent with the hypothesis that task variables are represented in a low-dimensional task-related subspace of the cortical activity space, and that attention increases the separation of the states in this task-related state space.

Supplementary Figures



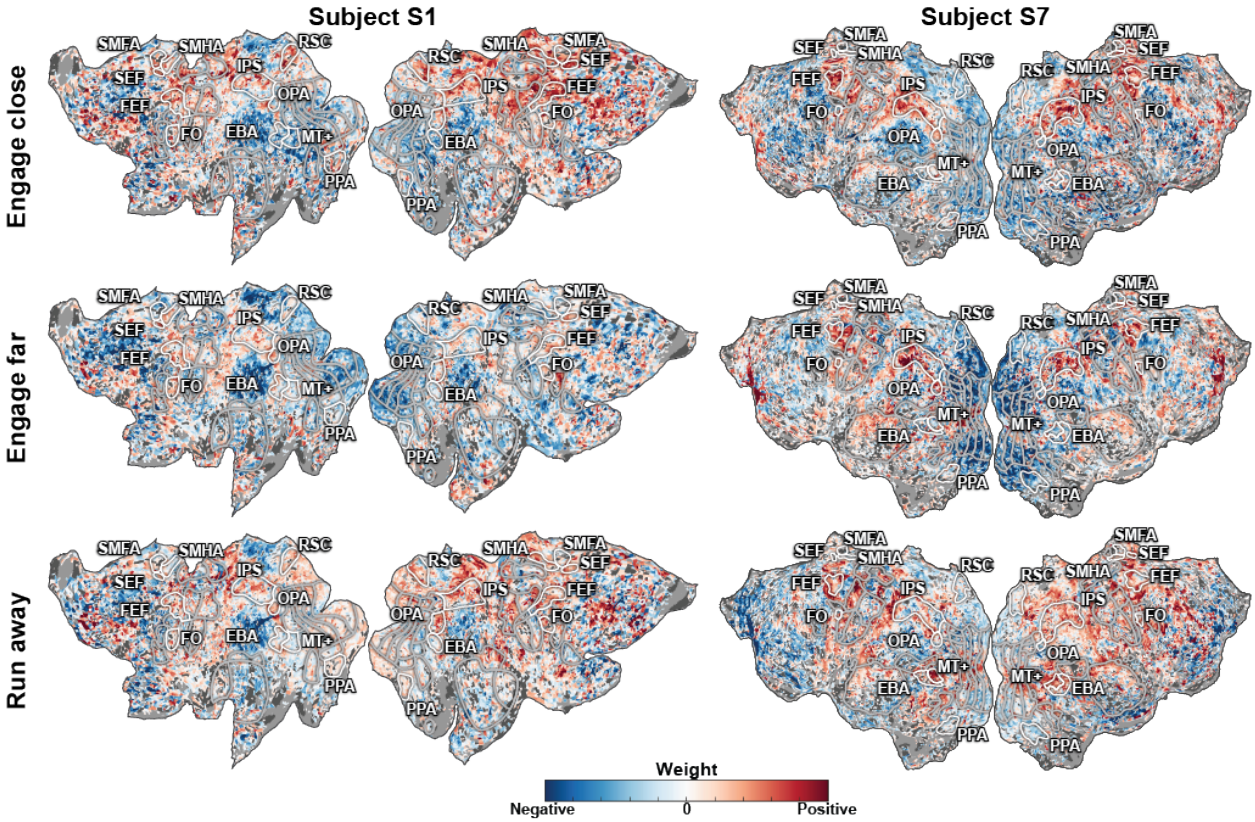
Supplementary Figures

Supplementary Figure 10. (Previous page) Task variables in the visual attention task are represented in broad functionally specialized networks distributed across the cerebral cortex. Model weights for human- and vehicle-presence are shown on the flattened cortical surface for subjects S2-S6. Blue corresponds to negative weights, white to no weights, and red to positive weights. Voxels are thresholded by the prediction performance of a semantic encoding model at a non-FDR-corrected $p < 0.05$ significance level. Selected ROIs are highlighted and labeled. These model weights agree well with the known functions of the cortex and further suggest that task variables are meaningfully represented in a low-dimensional subspace of cortical activity.



Supplementary Figure 11. Behavioral states in the video game task are represented in broad functionally specialized networks distributed across the cerebral cortex. Weights for "dead", "round start", "unsafe explore", and "flash" are shown on the flattened cortical surfaces for both subjects. Blue correspond to negative weights, white to no weights, and red to positive weights. Voxels are thresholded by the prediction performance of an encoding model at a Bonferroni-corrected $p < 0.05$ significance level. Selected ROIs are highlighted and labeled. These dimensions share representation in the TPJ, precuneus, and prefrontal cortex. Note that there are no weights for "flash" for subject C2, because C2 did not experience any "flash" events during the video game task.

Supplementary Figures



Supplementary Figure 12. Behavioral states in the video game task are represented in broad functionally specialized networks distributed across the cerebral cortex. Weights for "engage close", "engage far", and "run away" are shown on the flattened cortical surfaces for both subjects. Blue correspond to negative weights, white to no weights, and red to positive weights. Voxels are thresholded by the prediction performance of an encoding model at a Bonferroni-corrected $p < 0.05$ significance level. Selected ROIs are highlighted and labeled. These dimensions share representation in the in the motor, pre-motor, and supplementary motors areas, and also in IPS, FEF, and SEF.