Partially Mixed Selectivity and Parietal Cortex

Thesis by

Carey Y. Zhang

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2018

(Defended May 17, 2018)

Carey Y. Zhang
ORCID: 0000-0001-9867-4510

Acknowledgements

I would like to thank my advisor, Richard Andersen, for granting me the opportunity to do this work, providing an excellent research environment, and giving wise advice and recommendations. I would also like to thank Tyson Aflalo. His detailed guidance and friendship were invaluable to my growth and experience in the lab. I would like to thank participant NS for her time and patience, providing rich and interesting data to study.

I would like to thank Vasileios Christopoulos, Sofia Sakellaridi, Spencer Kellis, Boris Revechkis, Christian Klaes, Michelle Armenta Salas, Luke Bashford, Sumner Norman, Matiar Jafari, HyeongChan Jo, and Srinivas Chivukula for enjoyable and productive discussions, Kelsie Pejsa for administrative assistance, and Viktor Shcherbatyuk for technical assistance.

Additional thanks to the National Institutes of Health, the Boswell Foundation, and the Tianqiao and Chrissy Chen Institute for Neuroscience for providing the funding that supported this work.

Dedication

I dedicate this thesis to my mother and father for all they have done to raise and support me, ultimately making this undertaking possible. Without your care, I would not be who or where I am today.

I also dedicate this thesis to my brother and sister. My experiences growing up together with you have also shaped who I am today.

Abstract

Brain-machine interfaces (BMIs) decode intention signals and other variables from the brain in order to control a computer, tablet, or prosthetic limb. In order to improve the technology, a better understanding of the representational mechanisms within the brain is necessary. Here we study how the anterior intraparietal area (AIP) of human posterior parietal cortex is able to represent many variables within a small patch of cortex. We record single unit activity using a 4 x 4 mm microelectrode array implanted in AIP of a human tetraplegic volunteer. Testing movements of different cognitive strategies, body parts, and body sides, we find that the neural population represents information in a high-dimensional way, termed “mixed selectivity”, with individual units coding for idiosyncratic combinations of variables. Furthermore, we find that the variables are not randomly mixed but exhibited “partially mixed selectivity” with certain variables more randomly mixed than others. Representations were “functionally segregated”, with representations of the hand and shoulder largely orthogonal despite the high degree of anatomical overlap; representations of body side and strategy were organized by body part. We also examine how the representations changed between BMI training and online BMI control. We find that the structure of the movement representations was preserved, with the different representations found during calibration maintained during online control. Finally, we study the sensory mirror system, a system that processes observed sensations similarly to experienced sensations. We once again find partially mixed selectivity and functional segregation by body parts, showing that this method of encoding information exists not just in the action intention domain but also in the sensory domain. Our results propose partially mixed selectivity as a general mechanism for encoding high dimensional in formation in a small neural population, while also advancing the possibility of limited electrode-array BMIs decoding movements of a large extent of the body.

Published Content and Contributions

Zhang, C.Y. et. al. (2017). Partially Mixed Selectivity in Human Posterior Parietal Association Cortex. Neuron 95, 697-708. doi: 10.1016/j.neuron.2017.06.040. url: http://www.cell.com/neuron/abstract/S0896-6273(17)30592-5

C.Y.Z. participated in the conception, analysis, and interpretation of the data as well as in the writing of the manuscript.

Table of Contents

List of Illustrations and/or Tables

Supplemental Illustrations

Nomenclature

AIP: Anterior intraparietal area.

Anatomical segregation: Division of the population based on physical neurons/units.

Attempted movements: Movements where the participant actually tries to engage the muscles (regardless of physical success).

Baseline: A rest/relaxation state used to differentiate neural activity at rest from neural activity.

BA: Brodmann Area.

BMI: Brain-machine interface.

Closed-loop/Online-control: When there is feedback of control performance to the participant in real-time.

Decode: Converting input representations (typically neural signals) into some output, e.g., using a linear classifier or a linear model.

Effector: A body part or limb used to control an object (e.g., cursor) in a BMI environment.

Firing rate: The rate at which a unit is releasing spikes/action potentials. Firing rate codes information.

Functional segregation: Division of the population based on information they encode, rather than based on physical neurons.

Imagined movements: Movements where the participant is only visualizing them, without trying to engage any muscles.

IPL/IPS: Intraparietal region, around the intraparietal sulcus

LFP: Local field potential.

LIP: Lateral intraparietal area.

Motor mirror neurons: Neurons that process the observation of someone performing an similarly to oneself performing the action.

NHP: Non-human primate.

N.S.: The human participant/volunteer taking part in the study and the participant being studied in all experiments in this dissertation.

Open-loop: When there is no feedback of control performance to the participant.

Online: When information is being processed live, under real-time conditions.

Offline: When information is not being processed in real-time (e.g., post-hoc analysis).

Population: A group of units; considering the units in aggregate as opposed to as single units/individually.

PFC: Prefrontal cortex.

PPC: Posterior parietal cortex.

PRR: Parietal reach region.

SCI: Spinal cord injury.

SEM: Standard error of the mean.

Sensory mirror nrurons: Neurons that process the observation of someone feeling a sensation similarly to oneself feeling the sensation.

Spikes: Action potentials from when a neuron/unit fires.

Training: Initial data collection used to calibrate a decoder for BMI control.

Tuned: Being differentially responsive to one condition compared to others and/or baseline.

Unit: A neuron, a basic unit of computation within the brain.

1 Introduction

An estimated 12,000 new cases of spinal cord injury (SCI) occur each year in the United States, with over 50% of them leading to partial or complete tetraplegia (Foundation for Spinal Cord Injury Prevention, 2009). Tetraplegia involves the paralysis of all four limbs and the loss of sensation below the level of injury. As a result, it is difficult for tetraplegics to perform many everyday actions on their own. Neuroprosthetics and brain-machine interfaces (BMIs) offer a way to restore some ability and independence.

In recent years, there has been significant progress in the field of BMIs. BMI recordings from the brain have been able to successfully control robotic arms (Balasubramanian et al., 2017; Hochberg et al., 2012; Klaes et al., 2015; Meng et al., 2016), virtual computer cursors (Aflalo et al., 2015; Jarosiewicz et al., 2015), and even a patient's own muscles (Ajiboye et al., 2017). These BMIs generally work by having the patients or non-human primates (NHPs) imagine or attempt making movements in order to control a device.

Despite these engineering advances, however, there is a lack of understanding how the brain represents the many variables relevant to BMI cortical control and cognitive processes (e.g., different effectors, cognitive strategies, sequences of movements, attention, memory, sensations, etc.) as well as how these representations change after paralysis. In this dissertation, we are particularly interested in how the posterior parietal cortex (PPC) represents variables with respect to the different body parts being used for control, the relationships between the representations of the different movement types, and how the relationships change when used

for actual BMI control. Furthermore, we are interested in whether variables in other domains (e.g., the sensory domain) are represented in a similar framework.

In Chapter 2, we assess the different types of BMIs and the current state of the field. We discuss the PPC, the main brain area studied in this dissertation, in terms of traditional and more modern views of its function and organization, as well as the potential benefits of designing BMIs that record from PPC compared to primary motor cortex (M1), another brain area often used for BMIs. We also explain ways in which the brain can encode information, and in particular "mixed selectivity", an increasingly studied way of representing high-dimensional information in a relatively small network of neurons. Lastly, we review the literature on mirror neurons, which are neurons that appear to represent someone else's movements/sensations similarly to when making the movements/feeling the sensations themselves, effectively "mirroring" them.

In Chapter 3, we study how PPC encodes movements of different body parts on both the left and right side of the body when using different cognitive strategies. In particular, we find a high degree of anatomical overlap, with single units coding for idiosyncratic combinations of the tested variables, exhibiting the mixed selectivity discussed above. Furthermore, we find that the representations of body parts are largely orthogonal, "functionally segregating" the effector responses. Body side and strategy, however, were not coded in a mixed manner, as their representations were organized by body part. This type of "partially mixed coding" suggests that the way variables are encoded depends on the compared dimensions and is potentially

advantageous for neuroprosthetics, enabling a single array (recording from a small patch of cortex) to decode motor intentions from a large extent of the body.

In Chapter 4, we study the diverse movements during closed-loop BMI control, specifically focusing on how well the structure of the representations observed in Chapter 3 (during "training") are preserved during online control. Focusing on imagined/attempted left/right hand movements, we find that the representations of the different movement conditions are maintained equally well, with the structure consistent between training and online control. All the tested movement conditions were feasible for BMI control, with attempted movements of the right hand resulting in the best control performance. Furthermore, as a result of the maintained structure of the representations, we found that performance differences were predictable by data collected during training. In other words, the consistency of the structure made estimates of tuning and relative performance based on the training data meaningful during online control.

In Chapter 5, we study sensory mirroring responses in AIP and how they are represented at both a single unit and population level. Sensory mirror neurons are neurons that process observed sensations similarly to experienced sensations when the sensations are alike. Although not directly relevant to BMIs, the results of this chapter shed light on the generalizability of the above representational framework to other contexts (i.e., sensory/social interaction). Comparing the neural representations of felt and observed touches of the cheek and shoulder, we find single unit sensory mirror neurons. Furthermore, we find a population sensory mirroring response, with neural representations of felt and observed sensations matched by body part

(e.g., felt and observed cheek touches) more similar than neural representations of body parts matched by person (e.g., felt cheek and shoulder touches). In the population response, we once again find partially mixed selectivity, with the representation of who was touched functionally segregated by body part. We also find that the mirroring responses are present regardless of whether sensations are presented in video or live action, and whether sensations are observed with free gaze or with required fixation.

Finally, in Chapter 6, we summarize our findings and discuss potential future extensions of our work. Specifically, we make predictions on how the results of our study might appear in other brain areas, how functional segregation might be a novel way of defining a high-level brain area's primary function, and how our results might inspire novel methods in the field of BMIs.

2 Background

2.1 Brain-Machine Interfaces

A brain-machine interfaces (BMI), also often called a brain-computer interface, is a technology that allows for communication between the brain and an external device, such as a prosthetic limb or a computer. This communication can be unidirectional from brain to machine (as in the case of decoding motor intentions to control a robotic arm), unidirectional from machine to brain (as in the case of stimulation of the brain to evoke a sensation), or bidirectional (as in the case of a fully closed-loop prosthetic limb that responds to user intentions and also sends sensory feedback signals). BMIs measure brain activity either directly through neurons (the individual computational units of the brain) or indirectly through some proxy or aggregate measure (e.g., electroencephalography, local field potentials, functional magnetic resonance imaging, etc.).

Traditionally, BMIs have been developed in clinical contexts, with applications to help paralyzed patients or amputees. For example, BMIs have helped patients with “locked-in” syndrome communicate again (Chaudhary et al., 2015). Similarly, BMIs have helped tetraplegics perform some independent movements, either through control of a robotic limb (Aflalo et al., 2015) or through direct stimulation of their muscles to perform the movement (Ajiboye et al., 2017). More recently, there has even been interest in BMIs within the industry and the tech community. Several venture capitalist-backed startup companies, such as Kernel and Neuralink, and even some existing tech giants like Facebook, have started to explore BMIs in the past few years.

Typically, BMIs are driven by the user's motor intentions, often in the form of motor imagery (e.g., imagining arm reaches to different targets on a screen). In general, a BMI first learns the brain's neural representations corresponding to the different movement intentions in the absence of any actual control. This data is collected in a calibration ("training") phase either by having the user follow instructed movement cues or by having the user mimic the actions necessary to control the effector (e.g., following a mouse cursor's movements with their arm, with the cursor following some computer-defined trajectory). A neural decoder (e.g., a linear decoder) is then trained on this data to map the neural activity to the movement intentions in a supervised learning fashion. This trained decoder can then be used for online BMI control (Hochberg et al., 2006; Kim et al., 2008; Simeral et al., 2011).

There are several brain regions often used for BMI control. Primary motor cortex (M1) is one of those regions studied for BMI control. It is very involved in low-level motor output and as a result, many of the neural signals are strongly correlated with desirable motor control variables (e.g., degrees of freedom, trajectories, velocities, forces, etc.) (Georgopoulos et al., 1982; Hochberg et al., 2012; Holdefer and Miller, 2002; Morrow and Miller, 2003; Sergio et al., 2005). It is highly specialized, with specific, minimally overlapping networks for different parts of the body, as described in early studies detailing Penfield's homunculus (Lotze et al., 2000; Penfield and Boldrey, 1937). In other words, a small area of M1 is likely to have dense representations of only one body part.

In contrast to this specialized region, the posterior parietal cortex (PPC) has more overlapping networks and higher-level intention tuning, particularly in the regions around the intraparietal

sulcus (IPL), serving as a connection between sensory and motor areas of the brain (Andersen and Cui, 2009; Baldauf et al., 2008; Gail and Andersen, 2006; Ishida et al., 2010; Quiroga et al., 2006). (More information on PPC below). This higher-level intention-coding property provides an opportunity to decode the subject's intended set of actions and to act on them accordingly with intelligent assistive decoders, offloading a substantial portion of the burden of fine-grained control from the subject to the computer (Andersen et al., 2014; Musallam et al., 2004).

Regardless of the brain area, the neural activity that BMIs use is encoded in the activity of neurons. Neurons fire electrical signals (action potentials, "spikes") to signal information. The firing patterns of the neurons encode the information, either in the temporal (Singh and Levy, 2017; Stein et al., 2005) or frequency domains (Adrian and Zotterman, 1926; Majaj et al., 2015). This firing activity can be observed in a variety of ways. In functional magnetic resonance imaging (fMRI), the prevailing method is to monitor blood oxygen levels in the brain as a proxy for brain activity. This allows for specific regions to be associated with certain brain functions, such as facial recognition (Liu et al., 2010), internal thought (Benedek et al., 2016), semantic memory (Sugarman et al., 2012), motor imagery (Filimon et al., 2015; Pilgramm et al., 2016), and more. Another technique is electroencephalography (EEG). This technique records the brain's electrical activity from the scalp, using an array of electrodes to measure the gross activity of a region. EEG has been used to diagnose neurological disorders such as epilepsy (Noachtar and Remi, 2009) as well as to drive some BMIs (Bhagat et al., 2016; Höhne et al., 2014). Both fMRI and EEG are non-invasive techniques and typically do not require any surgical intervention to use.

Electrocorticography (ECoG) and intracortical recording, on the other hand, are invasive techniques and record the brain from beneath the skull directly. Like EEG, ECoG also records aggregate voltage signals from the brain. However, because the neural signals do not need to be conducted through the skull in ECoG, the spatial resolution is much higher, up to 5 mm^2 compared to 10cm^2 with EEG (Buzsáki et al., 2012). In intracortical recording, microelectrode arrays (MEAs) are implanted inside the gray matter of the brain. MEAs are not only able to record aggregate neural activity in the form of local field potentials (LFPs), but are also able to record the activity of single neurons in the forms of action potentials (spikes). Recent studies have also been able to develop BMIs using signals recorded from ECoG arrays (Leuthardt et al., 2004; Schalk et al., 2007) and MEAs, in both non-human primates (Chestek et al., 2011; Nuyujukian et al., 2017) and humans (Aflalo et al., 2015; Collinger et al., 2013; Hochberg et al., 2006; Jarosiewicz et al., 2015; Pandarinath et al., 2015; Taylor et al., 2002).

One major limitation, in part due to the invasiveness of the ECoG and MEAs, is that they can only record from a relatively small patch of cortex at a time. Compared to EEG which can cover most of the scalp, for example, the Neuroport array (an MEA from Blackrock Microsystems), records from only a 4 x 4 mm patch of cortex. MEAs may have higher spatial resolution, but they are gated by the amount of cortex they can cover. The limitation in the coverage may limit the variety of variables that can be decoded (e.g., the number of different body parts). Thus, there is potentially a tradeoff between the invasiveness of the arrays and the variety of information that can be recorded.

2.2 Posterior Parietal Cortex (PPC)

The posterior parietal cortex is a region of the brain historically implicated in a variety of higher-level brain functions. Having connections to both sensory and motor regions of cortex, PPC is considered an association area that processes sensory signals into more complex functions useful for motor functions and decision making, such as awareness, attention, and action planning (Balint, 1909; Holmes, 1918; Mountcastle, 1975; Ungerleider and Mishkin, 1982).

More recent studies in non-human primates have found regions of specialization around the intraparietal sulcus (IPS) within PPC, with different anatomical regions specializing in different body parts and functions. For example, the lateral intraparietal area (LIP) has been linked to saccades while the medial intraparietal area (MIP) has been found to be a part of the parietal reach region (PRR) and responsible more for reach planning (Andersen et al., 1987; Christopoulos et al., 2015; Quiroga et al., 2006). In the ventral intraparietal area (VIP), studies have found evidence of visuotactile mirroring, with units representing both observed and experienced tactile sensations (Ishida et al., 2010). Studies have found the anterior intraparietal area (AIP) to be selective for hand shapes (i.e., grasp types), hand movement intentions, and visual features of objects relevant for grasping (Klaes et al., 2015; Murata et al., 2000; Schaffelhofer et al., 2015; Schaffelhofer and Scherberger, 2016).

Inactivation studies in different regions on IPS also support the idea of anatomical specialization, with the inactivation of PRR leading to reach deficits (Christopoulos et al., 2015; Hwang et al.,

2012), the inactivation of LIP leading to saccade deficits (Li and Andersen, 2001), and the inactivation of AIP leading to grasp deficits (Gallese et al., 1994). At the same time, however, these areas are also highly interconnected (Andersen et al., 1990) with some inactivation studies finding deficits with coordinated movements. For example, inactivation of PRR was found to cause deficits in coordinated simultaneous reaches and saccades but not saccades alone (Hwang et al., 2014).

In light of the heavy interconnectivity, it is reasonable to expect significant crosstalk between the regions around IPS and thus a high degree of overlap between representations in IPS. Indeed, there have been some studies finding evidence of these functions being more anatomically overlapping, with the representations of different body parts mixing (Astafiev, 2003; Connolly et al., 2003; Culham, 2003; Gallivan et al., 2011; Prado, 2005) and the representations of body parts on different sides of the body as well (Gallivan, 2013).

Despite the mixed picture on the degree to which the regions around the IPS are specialized by movement intention type, it is clear that said regions encode a rich diversity of information. Among the neurons in PPC, some encode information about the visual scene (Ishida et al., 2010; Murata et al., 2000), while others encode information about movement intentions and overall goals (Andersen and Buneo, 2002). These movement intention signals are present during both execution and movement planning, even encoding an entire sequence of individual movements ahead of time (Baldauf et al., 2008).

Furthermore, past studies in PPC on gain fields have suggested that individual neurons in PPC can be tuned to multiple variables simultaneously. Specifically, the gain field studies focused on how different spatial reference frames encoded at the neural level. They found that neurons in and around LIP encode the spatial location of objects relative to head, eye, and body position, allowing for a stable representation of the outside world in body-centered coordinates (Brotchie et al., 1995; Zipser and Andersen, 1988).

More recently, the idea of PPC neurons being tuned to multiple variables has been extended past references frames and representations of space towards other, more categorical, variables in a structure known as “mixed selectivity” (Raposo et al., 2014). In mixed selectivity, individual neurons can be simultaneously tuned to various idiosyncratic combinations of variables. Mixed selectivity allows for a relatively small network of neurons to encode more information and in higher dimensions than traditional “pure selectivity”, where individual units are tuned to only one variable at a time (Fusi et al., 2016). Mixed selectivity has also been found in other association cortices of the brain such as prefrontal cortex (Rigotti et al., 2013), and is seen as important for a brain area to perform high-level computations and functions (see Mixed Coding for more details).

In the context of BMIs, the coding of many high-level variables in PPC offers an alternative brain area to M1 with unique benefits for BMI implementation and efficiency. The ability to decode the user’s high-level goal from PPC enables a BMI to use computer assistance (e.g., artificial intelligence) to automatically perform the necessary steps to successfully accomplish the goal, such as with smart robotics. In contrast, a BMI depending on only the trajectory or velocity

information encoded in a brain region (such as in M1) would require the user to carefully perform each of the necessary steps on their own. In Aflalo et al. (2015), for example, they found a unit in PPC activated specifically when the participant imagines moving his hand to his mouth, but not for any movement with a similar trajectory (e.g., hand to shoulder or hand to ear). In this case, one can imagine a BMI using the signal to trigger a robotic limb to move to the user's mouth automatically, as opposed to requiring the user to carefully control each joint and degree of freedom simultaneously.

Furthermore, the mixed selectivity coding structure suggests that it is possible for a BMI recording from a small patch of PPC to decode many different variables and body parts. This has an obvious advantage of allowing a single array to record from a large extent of the body, such as the hands and arms (Aflalo et al., 2015; Klaes et al., 2015) or bilaterally (Chang and Snyder, 2012). M1, in comparison, usually only represents limbs on the contralateral side of the body and in a less overlapping way (Fritsch and Hitzig, 1960; Lotze et al., 2000). As imagined body part movements are often used to control BMIs, the ability to decode movements of multiple different body parts naturally increases the control possibilities and degrees of freedom.

The ability to decode more variables and body parts from a small patch of PPC also provides a benefit from a patient safety perspective. In terms of recording hardware, it is desirable to record at a high spatial resolution (i.e., single units) while also minimizing the invasiveness of the recording technology. Currently, MEAs implanted invasively in the brain are the prevailing method for recording at a high enough spatial resolution in humans. By recording from a smaller patch of cortex, the craniotomy required to implant the MEA would also be smaller, resulting in

a relatively safer procedure with fewer risks (Cho et al., 2017; Regan et al., 2015). Thus, the potential to record a large amount of information from a small patch of PPC using an MEA is attractive from a clinical safety perspective, as well.

2.3 Mixed Coding

Mixed coding (or "mixed selectivity") is a mechanism for how neural populations represent information in the brain. Traditionally, studies have either looked at neural representations from a "pure selectivity" perspective, where individual neurons are tuned only to single task variables, or a "sparse selectivity" perspective, where neurons are tuned to specific combinations of task variables. Mixed selectivity, on the other hand, is a framework where the neurons in a population are tuned, often nonlinearly, to idiosyncratic combinations of variables, i.e., neither completely pure nor sparse but rather with some units demonstrating examples of each.

In 1988, Zipser and Andersen studied how neurons in PPC near LIP behave during a saccade, particularly focusing on how they encode both the position of the eye and the target of the saccade (Zipser and Andersen, 1988). They found that the neurons were multiplicatively tuned to both eye position and target location, an encoding structure termed "gain fields". Despite the mixing of eye and target positions, however, the population as a whole could still be used to predict eye position and target locations. Other studies have also found examples of gain fields in various parts of the brain, such as the mixing of visual and tactile receptive fields in VIP (Avillac et al., 2005) or the mixing of hand, eye, and goal positions during reach planning in dorsal premotor cortex (PMd) .

These studies on gain fields can all be considered early work in the field of mixed selectivity in the context of spatial representations, with neurons in the population tuned to various combinations of receptive fields. More recently, however, the concept of mixed selectivity has been extended to a more diverse variety of contexts, such as in the context of categorical decision making (Raposo et al., 2014; Rigotti et al., 2013).

In these studies, mixed selectivity has been found in associative areas of the brain such as PMd, PFC, and PPC and is seen as an important hallmark of higher level cognitive areas or areas related to associative learning (Pesaran et al., 2006, 2010; Raposo et al., 2014; Rigotti et al., 2013). Studies on gain fields, for example, interpret the mixing of different receptive fields as necessary in the transformation of variables from one reference frame to another, a process important in transforming high level intentions into actions (Pesaran et al., 2006, 2010; Salinas and Sejnowski, 2001; Salinas and Thier, 2000). A recent modeling study found that the degree of mixed selectivity observed in NHP PFC was consistent with models of Hebbian learning (Lindsay et al., 2017), where the connection between two neurons strengthens when they have highly correlated outputs (Morris, 1999). This consistency suggests a relationship between mixed selectivity in association areas and training over time.

The mixed selectivity framework offers several advantages over the more traditional pure or sparse selectivity frameworks. A sparse selective representation faces the problem of exponential blowup. In sparse selectivity, each neuron is tuned to a specific combination of task variables. As the number of variables increases, however, the number of neurons required to

encode that information increases exponentially. This is a broader phenomena from machine learning, computer science, and statistics known as the “curse of dimensionality”. In general, the curse of dimensionality describes how as the number of dimensions in the data increases, the amount of space/samples required to fully represent the possible permutations of the features increases exponentially (Bellman, 2015).

In the context of neural representations and sparse selectivity, it means that not only does the number of neurons exponentially increases with the number of variables, but the number of connections from these neurons to downstream areas processing the information explodes exponentially as well (Anderson and Kreiman, 2011; Fusi et al., 2016). This is clearly not a feasible method of representation for all the variables encoded in the brain. However, this problem is largely reduced with mixed selectivity. The presence of neurons tuned to various combinations of variables allows for the neurons to represent more than a single variable, allowing them to be useful for a much larger variety of applications and downstream applications. This flexibility of use also allows for the readout of the variables to require much fewer connections downstream (Fusi et al., 2016; Pouget and Sejnowski, 1997).

Compared to pure selectivity, where units are tuned to individual variables, mixed selectivity allows for information to be encoded in higher dimensions (Rigotti et al., 2013). The higher dimensionality allows for linear readout of many variables that would be otherwise inseparable in lower dimensions (e.g., with pure selectivity), a concept well understood in machine learning (Fusi et al., 2016). Linear readout is easily implemented by neurons and has been found to be sufficient to solve many complex tasks, such as predicting chaotic systems and spatiotemporal

computations (Adibi et al., 2014; Buonomano and Maass, 2009; Jaeger and Haas, 2004; Shamir and Sompolinsky, 2006). Thus, the ability enabled by mixed selectivity to decode variables via simple linear readout is highly desirable. Note, however, that for the mixed selectivity to encode variables in higher dimensions the mixing must be nonlinear as opposed to linear. Purely linear mixing of variables would fail to create any additional separability in the neural representations (i.e., not increase the dimensionality of the representation) (Fusi et al., 2016; Rigotti et al., 2013), potentially making the aforementioned simple linear readout of all variables impossible.

So far, studies on mixed selectivity have looked at behaving rodents and NHPs, with results suggesting that all task variables are randomly mixed together (Pesaran et al., 2006, 2010; Raposo et al., 2014; Rigotti et al., 2013). However, it is unclear whether this is indeed the case. The random mixing of apparently all variables could be due to the limited number of task variables tested in an experiment, for example. The random mixing also raises questions about how different brain areas can have specialization of function while remaining randomly mixed. For example, studies in PPC have found areas of specialization to saccades, reaching movements, grasps, etc. (see Posterior Parietal Cortex above for more details), but it is unclear how these areas would be different in the framework of mixed selectivity.

Past studies have failed to study a wide enough set of task variables to address these questions. Studies have primarily used laboratory animal models, and animals are typically difficult to train to perform many tasks with many variables. This limit on the number of simultaneously testable task variables makes it relatively problematic to study high dimensional mixed selectivity in animals. Humans, on the other hand, can be instructed to perform tasks involving multiple

variables easily, with little to no overhead when testing new task variables. However, there have so far been no studies recording from high-level associative areas in human cortex in the context of mixed selectivity.

2.4 Mirror Neurons

About 25 years ago, researchers at the University of Parma in Italy discovered neurons in the premotor cortex of macaque monkeys responsive when observing someone else perform an action as well as when the monkey was performing the same action itself (Di Pellegrino et al., 1992; Gallese et al., 1996). The discovery of these "mirror neurons" created a field studying how mirror responses might relate to larger behaviors in social interaction.

The majority of the work in this field has been in NHPs using both fMRI and electrophysiology (Caggiano et al., 2009; Filimon et al., 2007; Fogassi et al., 2005; Fujii et al., 2008; Grèzes et al., 2003; Keysers et al., 2003; Rozzi et al., 2008). These studies have mostly found mirror neurons in brain regions involved in action planning such as premotor cortex and PPC.

More recently, there have been some mirror neurons studies in humans, too. These studies have used fMRI, recording brain areas in response to observed actions and imagined/executed movements. Consistent with the NHP studies, they have found mirror responses in premotor cortex, inferior frontal gyrus, and the IPS of PPC (Chong et al., 2008; Kilner et al., 2009; Tai et al., 2004). A recent electrophysiology study in humans also found single unit examples of motor mirror neurons in the supplementary motor area (SMA) (Mukamel et al., 2010).

The prevailing school of thought is that these mirror neurons are related to recognizing and understanding the actions and intentions of others, allowing people to imitate/mimic each other, and ultimately forming a basis for interactive social behaviors (Caggiano et al., 2011; Caramazza et al., 2014; Fogassi et al., 2005; Iacoboni, 2009; Keysers and Gazzola, 2006; Rizzolatti and Fabbri-Destro, 2008). Some even argue that mirror neurons are evidence of the biological hardwiring of empathy into animal/human behavior (Gazzola et al., 2006; Iacoboni, 2009); empathy that is critical for the formation of societies, altruistic behaviors, and morals (Adolphs, 2009; Tangney et al., 2007; Waal, 2008).

At the same time, however, some maintain that the motor mirror system is not the substrate for action understanding. They argue that there has not been any direct tests of the mirror neurons actually representing action understanding and that action understanding could be achieved through other higher level areas (Hickok, 2009). They also argue that mirror neurons could just be a natural emergent phenomena of associative learning over time, rather than a biologically-directed feature (Heyes, 2010).

The majority of the work on mirror neurons has focused on “motor mirror neurons”, neurons tuned to both the performance of an action and the observation of someone else performing the same action. There is in fact a second, less well-studied, class of mirror neurons, termed “sensory mirror neurons”. Analogous to motor mirror neurons, sensory mirror neurons are responsive when feeling a sensation as well as when observing someone else feel the same sensation. More precisely, sensory mirror neurons are neurons that process the observation of

someone else feeling a sensation similarly to how they process the actual experience of the sensation.

Single unit studies in NHPs have found evidence of tactile sensory mirroring in areas around the IPL in PPC (Ishida et al., 2010). In humans, fMRI studies have also found the IPL and the secondary somatosensory cortices to respond similarly for observed and felt tactile sensations and even pain (Keysers et al., 2004; Osborn and Derbyshire, 2010).

In several studies, there have been cases of people with mirror-touch synesthesia, people who report being able to actually feel a sensation just from observing it (Banissy et al., 2009). fMRI studies on these people revealed hyperactivity in brain regions where sensory mirror neurons have been found (Blakemore et al., 2005). Subsequent studies have found that people with mirror-touch synesthesia tend to have higher levels of empathy, with some people even arguing that mirror-touch synesthesia could be linked to the ability to distinguish the identity of self from others (Banissy and Ward, 2013; Banissy and Ward, 2007). These studies further together suggest that the mirror system, and in particular the sensory side of the mirror system, could be related to empathy.

To the best of our knowledge, however, there have so far been no electrophysiology studies on the sensory mirror system in humans. Without single unit recordings, the structure of how these mirror systems manifest is unclear. Although able to cover a large extent of the brain, fMRI studies are generally limited in their spatial resolution (Glover, 2011). Recordings of single units,

on the other hand, would be able to answer questions about the structure of the sensory mirror responses (e.g., how much the subpopulations tuned to felt and observed sensations overlap, potential tuning differences between felt and observed sensations, etc.).

3 Partially Mixed Representations in Posterior Parietal Cortex

The following chapter's contents are taken and adapted from Zhang et al. 2017, with modifications done to fit the dissertation format.

Zhang, C.Y. et. al. (2017). Partially Mixed Selectivity in Human Posterior Parietal Association Cortex. Neuron 95, 697-708. doi: 10.1016/j.neuron.2017.06.040.

3.1 Introduction

The posterior parietal cortex (PPC) of humans has historically been viewed as an association area that receives diverse inputs from sensory cortex, "associates" these inputs for processing more cognitive functions such as spatial awareness, attention and action planning, and delivers the outcomes of the associative process to more motor regions of the frontal cortex (Balint, 1909; Holmes, 1918; Mountcastle, 1975; Ungerleider and Mishkin, 1982). However, subsequent single neuron recording experiments with behaving non-human primates (NHPs) point to a systematic organization of functions in PPC (Andersen and Buneo, 2002). Of particular interest to the current investigation, separate cortical areas around the intraparietal sulcus (IPS) have concentrations of neurons selective for saccades (lateral intraparietal area, LIP) (Andersen et al., 1987), reach (parietal reach region, PRR) (Snyder et al., 1997) and grasping (anterior intraparietal area, AIP) (Murata, 2000). These data suggest that this part of the PPC, rather than being one large association region, is rather composed of a number of anatomically separated cortical fields that are specialized for intended movements that are effector-specific (eye, arm, hand).

More recent functional magnetic resonance imaging (fMRI) studies in humans have presented a mixed picture with some studies finding similar segregation for the types of intended movement in areas around the IPS (Astafiev, 2003; Connolly et al., 2003; Culham, 2003; Gallivan et al., 2011; Prado, 2005) and other studies finding largely an intermixing of effectors (Beurze, 2009; Heed et al., 2011a; Hinkley, 2009; Levy, 2007) as well as bimanual representation (Gallivan, 2013). These findings provide evidence for a degree of distributed and overlapping representation of effectors on both sides of the body within PPC.

With the first chronic single neuron recordings of PPC in humans, we found similarities with the NHP studies. Neurons in human AIP are highly selective for different imagined grasp shapes while neurons in nearby Brodmann area (BA) 5 are not (Klaes, 2015). However, the human neural recordings also pointed to some degree of distributed representation, with AIP neurons also selective for reach direction and with AIP and BA5 neurons being selective for reaches with either the left or the right limb or both (Aflalo et al., 2015). While we have found evidence that multiple effectors are encoded in the same anatomical region of cortex, these studies were carried out in separate sessions and thus the functional organization of multiple effectors within the same population of neurons remains unclear.

Pertinent to how different effectors are coded within PPC are recent results that address encoding strategies for multiple dimensions of representations and their computational advantages in association cortices more generally. Neurons in prefrontal cortex and PPC (Raposo et al., 2014; Rigotti et al., 2013) exhibit what has been termed mixed selectivity (Fusi et al.,

2016), a neural encoding scheme in which different task variables and behavioral choices are combined indiscriminately in a non-linear fashion within the same population of neurons. This scheme generates a high-dimensional non-linear representational code that allows for a simple linear readout of multiple variables from the same network of neurons (Fusi et al., 2016). A basic question is whether such an organization of functional variables is universal or, rather, is in part due to the types of functional variables that were compared or the cortical subregions selected for study.

In the current study, we examine the anatomical and functional organization of different types of motor variables within a 4 x 4 mm patch of human AIP. We varied movements along three dimensions: the body part used to perform the movement (hand versus shoulder), the body side (ipsilateral versus contralateral), and the cognitive strategy (attempted versus imagined movements). Each of these variables has been shown to modulate PPC activity (Andersen and Cui, 2009; Gallivan, 2013; Gerardin et al., 2000; Heed et al., 2011a). Thus we are able to look at how different dimensions of motor variables are encoded, and whether different variable types are treated in an equivalent manner (e.g., all variables exhibiting mixed-selectivity) or whether different functional organizations are found for different types of variables. Finally, we compare the hand and shoulder movements to speech movements, a very different type of motor behavior.

We find that movements of the hand and shoulder are well represented in human AIP, whether they are imagined or attempted, or performed with the right or left side. Single units were heterogeneous and coded for diverse conjunctions of different variables: there was no evidence

for specialized subpopulations of cells that selectively coded one movement type. However, the different motor dimensions were not indiscriminately mixed, as body side and cognitive strategy were fundamentally different from body part at the level of neural coding. There was a high-degree of correlation between movement representations of the right and left side, within, but not between body parts. The same was true for cognitive strategy. Thus, body part acted as a superordinate variable that determined the structure of how the other variables were encoded. Mixed-coding of some movement variables, but not others, argues in favor of PPC having a partially-mixed encoding strategy. Finally, while AIP lacks anatomical segregation of body parts, the mixed-coding between body parts leads to what we call *functional* segregation of body parts. Such segregation is hypothesized to enable multiple body parts to be coded in the same population with minimal interference.

3.2 Methods

3.2.1 Experimental Model and Subject Details

Subject N.S. is a 59-year-old female tetraplegic 7 years post-injury and has a C3-C4 spinal lesion (motor complete), having lost control and sensation in her hands but retaining movements and sensations in her upper trapezius. In this paper we refer to contraction of the upper trapezius as "shoulder movements" as short-hand for the resulting shoulder shrugging movement. The studies were approved by the California Institute of Technology, University of California, Los Angeles, and Casa Colina Centers for Rehabilitation Internal Review Boards. Informed consent was obtained from the participant N.S. after the nature of the study and possible risks were explained. Study sessions occurred at Casa Colinas Centers for Rehabilitation.

3.2.2 Behavioral setup

All tasks were performed with N.S. seated in her motorized wheel chair. Tasks were displayed on a 27-inch LCD monitor in a lit room. The monitor was positioned so that the screen occupied approximately 40 degrees of visual angle. Stimulus presentation was controlled using the Psychophysics Toolbox (Brainard, 1997) for MATLAB. No eye fixation was required or enforced.

3.2.3 Physiological recordings

Subject N.S. was implanted with two 96-channel Neuroport arrays (Blackrock Microsystems model numbers 4382 and 4383) in putative homologues of area AIP and Brodmann's Area 5d. Array placement was determined based on preoperative fMRI (Aflalo et al., 2015) and the array was placed at Talairach coordinate [-36 lateral, 48 posterior, 53 superior]. Neural activity was amplified, digitized, and recorded with the Neuroport neural signal processor (NSP). The Neuroport System, comprising the arrays and NSP, has received FDA clearance for <30 days acute recordings. We received FDA IDE clearance (IDE #G120096, G120287) to extend the duration of the implant for the purposes of a brain-machine interface clinical study using signals from posterior parietal cortex.

During recording, thresholds for action potential detection were set at -4.5 times the root-mean-square after high pass filtering (250 Hz cut-off) the full-bandwidth signal sampled at 30 kHz in the Central software suite (Blackrock Microsystems). Each waveform was composed of 48 samples (1.6 ms) with 10 samples prior to triggering 38 samples after. Single and multiunit activity was sorted by k-mediods clustering using the gap criteria to estimate the total number of clusters (Tibshirani et al., 2001). Clustering was performed on the first n principal

components, where n was selected to account for 95% of waveform variance. Post-hoc review of sorted unit statistics showed that channels were sorted with between 2-4 principal components (see Figure S 2A). Results of offline sorting were reviewed and adjusted if deemed necessary following standard practice (Harris et al., 2016). Only neurons recorded from the array implanted in putative AIP were analyzed. Pooling across all versions of the task, on average 93 sorted units were recorded from N.S. per session. Furthermore, to avoid bias, all spike sorting was performed prior to any analysis and blind to a unit or channel's response during the task. We used several metrics to quantify sort quality (see Figure S 2B-F) including 1) the percentage of interspike intervals (ISIs) shorter than 3ms, 2) the signal-to-noise ratio (SNR) of the mean waveform, 3) the between spike projection distance (Pouzat et al., 2002), 4) the modified coefficient of variation of the ISI (CV2), and 5) the cluster isolation distance (Harris et al., 2000) of each sorted cluster.

We recorded electromyogram (EMG) activity over the right trapezius muscle using B&L Engineering EMG electrodes. Raw analog EMG activity was fed into the NSP, aligned with neural signals, and sampled at 2 kHz. Signals were band-pass filtered (5^{th} order Butterworth filter with cut-off frequencies of 10 and 250Hz), full-wave rectified, and smoothed (box-car, 50ms window).

3.2.4 Task procedure

Several versions of a delayed movement task were constructed to determine the extent of tuning to control strategy within the neural populations recorded from AIP. In the primary task (Figure 1A), N.S. was cued for 2.5 seconds to what strategy (imagine or attempt), side (left or

right), and body part (hand or shoulder) to use, e.g., attempting to squeeze the right hand. In total there were eight possible actions which were pseudorandomly interleaved such that each condition was performed once before repetition. After a delay of 1.5 seconds, N.S. was cued to perform the cued action. Between each trial there was a 3 second inter-trial interval (ITI). Hand movements were hand squeezes while shoulder movements were shoulder shrugs (contraction of the trapezius). We ran 64 trials (8 trials per condition) on each session. This task was run over the course of 4 non-consecutive days. In total 357 units were recorded across the four recording sessions. Unless otherwise indicated, all figures were generated from data collected from this version of the task.

In a separate set of sessions, we repeated the experiment with the modification that shoulder shrugging movements were replaced with shoulder abduction in the frontal plane. Attempted shoulder abduction resulted in no overt movement and thus allowed us to compare body part representations exclusively below the level of injury. Six sessions run over the course of 6 non-consecutive days were recorded resulting in 629 recorded units. Each session contained 64 trials (8 trials per condition).

Movements of the shoulder and hand are frequently made together during natural behavior. We modified the delayed movement task by adding “speak left” and “speak right” as two actions unrelated to any hand or shoulder movements. To avoid overly long data collection sessions (as determined by patient feedback), we minimized the number of conditions by splitting sessions into either hand or shoulder movements exclusively, resulting in 6 conditions pseudorandomly interleaved (Imagine Left, Imagine Right, Attempt Left, Attempt Right, Speak

Left, Speak Right). Three sessions were recorded for the hand and the shoulder separately, with each session containing 72 trials (12 trials for each condition). In total 299 units were recorded for sessions using the hand while 228 units were recorded for sessions using the shoulder.

3.2.5 Unit selection

Analyses were performed on all units regardless of sort quality for statistical power and ease of presentation. To ensure that such pooling did not bias the conclusions of this paper, we performed the analyses separately on well isolated versus potentially multi-unit activity and found the results to be similar (see Figure S 1 for more details and also Unit quality classification below for how high-quality single units were identified). Units were pooled across days assuming independent populations across recording days. Analysis of separate days was also performed to demonstrate stability of results across sessions (Figure S 5). Units with mean firing rates less than 1.5 Hz were excluded from the analysis so that low firing rate effects would be minimized.

3.2.6 Linear analysis 1

We used a linear regression analyses to quantify tuning to each condition (e.g., for experiment 1, 8 total conditions from the possible combinations of the 2 strategies, 2 body parts, and 2 body sides). We created a design matrix consisting of indicator variables for each condition (e.g., the indicator for right attempted shoulder movements would consist of a vector where data points associated with right attempted shoulder movements would be assigned a 1, while all other conditions and the baseline samples would be assigned a 0). We then estimated firing rate as a linear combination of these indicator variables:

$$FR = \sum_{c} \beta_c X_c + \beta_0$$

where FR is the firing rate, X_c is the vector indicator variable for condition c, β_c is the estimated scalar weighting coefficient for condition c, and β_0 is a constant offset term. In such a model, the estimated Beta coefficients represent the expected firing rate changes from baseline for each condition. Tuning to each condition (Fig. 2A) was based on the p-value of the t statistic for each associated beta coefficient. This definition of whether a unit is significantly tuned to a condition is used as an inclusion criterion for some analyses, with the significance level (e.g., $p < 0.05$, uncorrected or Bonferroni corrected) depending on the specific analysis. The significance level of the differences between the number of units tuned to each condition was calculated using a two-sided Wilcoxon rank sum test on the distribution of p-values for each pair of conditions (Figure S 3).

3.2.7 AUC analysis

We performed a ROC analysis to quantify tuning strength for each condition. For each unit, strength of tuning was summarized as the area under the curve (AUC) when comparing each condition's Go or Delay neural response to baseline. The AUC values can range from about 0 to 1, with 1 indicating that every go/delay measurement is **greater than** every baseline measurement (excitatory response) and 0 indicating that every go/delay measurement is **less than** every baseline measurement (inhibitory response). To summarize the population (Figure 2B), we combined excitatory and inhibitory responses by reversing condition labels for AUC responses below 0.5. The separated responses are reported in Figure S 4. Only units with significant AUC ($p < 0.05$, permutation test) are included in the population average and thus we answer the question "what is the average strength of tuning for each condition for units with

significant tuning to the condition." Thus, this measure is descriptive and not a statistical assessment of significant tuning in the population. Pair-wise differences between the AUC for each condition were calculated using a two-sided Wilcoxon rank sum test on the significant AUC values (Figure S 3).

3.2.8 Linear analysis 2

Above we performed a linear regression analyses where firing rate was modeled as a function of each condition response. Here we perform an additional linear analysis where firing rate is modeled as the linear combination indicator variables for each motor variable (strategy, body side, and body part) and their interaction. All temporal epochs were identical, however the design matrix was updated to reflect the new model:

$$FR = \beta_1 Strategy + \beta_2 BodySide + \beta_3 BodyPart + \beta_4 Strategy * BodySide + \beta_5 Strategy * BodyPart + \beta_6 BodySide * BodyPart + \beta_0$$

Each unit was classified as being tuned to a term if the p-value of the corresponding beta coefficient was significant (i.e., $p < 0.05$, uncorrected). To examine the effect of the different motor variables on firing rate patterns across the population we performed a MANOVA test on the linear beta coefficients of the model. All units were used in the test (regardless of whether they showed tuning to a variable or not).

3.2.9 Test of population bias in representing motor variables

To examine whether there were systematic biases for the three different motor variables (strategy, body side, body part) across the population we performed a MANOVA test. The baseline firing rate of each neuron (taken during the intertrial interval) was subtracted from the

firing rate of the neuron during the Go phase, and this baseline-subtracted firing rate was used in the test. All units were used in the test (regardless of whether they showed tuning to a variable or not).

In comparing between the 8 movement conditions and the speaking conditions, we performed a t-test on the baseline subtracted firing rates, with the 8 movement conditions pooled into one group and the 2 speaking conditions pooling into a second group. Once again, all units were used regardless of their tuning.

3.2.10 Degree of specificity

We used a degree of specificity analysis to begin to understand how the different movement attributes were encoded relative to each other in individual neurons. This allowed us to test for population level tendencies for e.g., exclusive activation for one effector versus the other or whether tuning tended to be overlapping. We computed the degree of specificity for each motor dimension (e.g., left vs right, imagine vs attempt, hand vs shoulder) as the normalized difference in beta values computed for each condition $((|\beta_1| - |\beta_2|)/(|\beta_1| + |\beta_2|)$. Beta value extraction is described above (see Linear Model 1). The degree of specificity ranges from -1 to 1. A value of 1 indicates exclusive modulation for comparison variable 1, a value of -1 indicates exclusive modulation for comparison variable 2, while a value of 0 indicates comparable activation to both variables. Degree of specificity was only computed if at least one beta coefficient in the comparison was significant ($p < 0.05$, uncorrected). We used a Wilcoxson signed rank test to determine whether the median of the distribution of specificity values was significantly shifted from 0.

3.2.11 Correlation between representations

We used a population correlation analysis to measure the similarity between the neural representations of each condition. For each neuron, the beta coefficient for each condition (see Linear Model 1) was normalized by the corresponding 95% confidence interval to ensure a common scale proportional to signal-to-noise. The normalized beta values for each unit and for each condition were used to create a vector summarizing the population response. We used the correlation between these vectors to measure the similarity in neural space between conditions. Only units with a significant beta coefficient for at least one condition ($p < 0.05$, Bonferroni corrected) were included in the analyses.

We used hierarchical clustering (agglomerative hierarchical cluster tree; using the built-in MATLAB 2016a linkage and dendrogram functions with unweighted average correlation as the measure) to summarize the structure in the patterns of correlation between the different conditions (McKenzie et al., 2014).

3.2.12 Decoder analysis

Functional segregation of body parts should result in minimal shared representations of other motor variables across different body parts. Conversely, the lack of functional segregation between different body sides (or strategy) should lead to comparatively greater shared representation of the other motor variables across different body sides (or strategies). To test this, we trained a linear classifier (linear discriminant with equal diagonal covariance matrices) to differentiate between the two levels of one motor variable, restricting the training data to

one level of a second motor variable and then applying the classifier to the other level. Cross-validated performance restricted to the first level was also computed as this provided an upper bound on classification accuracy given the signal to noise of our data. For instance, we trained a classifier to differentiate left from right movements on shoulder movement trials and tested it on hand movement trials, also computing cross-validated performance within shoulder trials (Figure 7C). For features, we used firing rates from the first 2 seconds of the "Go" phase for units with significant tuning to any of the eight movement conditions. Only units significantly tuned to at least one of the 8 movement conditions ($p < 0.05$, Bonferroni corrected) were included in the analysis. Classifier performance was determined to be above chance if its performance was greater than 95% of decoders trained on randomly shuffled data (1000 shuffles).

3.2.13 Condition classification

We performed classification analyses using linear discriminant analysis with equal diagonal covariance matrices for each condition. Classification accuracy was estimated using stratified leave-one out cross-validation. Classification features were constructed using the first 2 seconds of the "Go" phase. Only units with significant tuning to any of the eight movement conditions were used ($p < 0.05$, Bonferroni corrected). Significant units were estimated from the training data and applied to the test data for each fold of the cross-validation routine to avoid peaking effects. Units were pooled across days for analyses. For each fold (1000 repetitions), one feature per condition was randomly sampled as the test data, with all other samples used for training.

3.2.14 Unit quality classification

For the analysis in Figure S 1, we needed to separate spike sorted units into high-quality single units and multi-units. This classification was done based on the cluster isolation distance (Harris et al., 2000) with a threshold of 10^1.6 dividing high-quality single units and multi-units. This threshold was chosen based on visual inspection of the distribution of all cluster isolation distance values (see Figure S 2).

3.3 Results

Recording from AIP of a female, C3/C4 tetraplegic participant 7 years post-injury (N.S.), we compared neural responses of attempted and imagined actions of the hand or shoulder on the right and left side of the body. Hand movements involved squeezing the hand into a fist and shoulder movements involved shrugging the shoulder. Shoulder shrugs are a staple of the participant's behavioral repertoire being a primary method to operate her motorized wheelchair. For imagined movements, we instructed N.S. to visualize her limb performing the instructed action, while for attempted movements, she was instructed to send the appropriate motor command to move the instructed limb. In the case of shoulder movements, attempted movement resulted in overt motor execution, while for the hand, there was no resulting movement because of paralysis. For the shoulder, we confirmed behavioral compliance by measuring the presence of trapezius EMG activity during attempted but not imagined movement.

We used a delayed movement paradigm (Figure 3.1A). Following an inter-trial interval (ITI), N.S. was instructed to attempt or imagine movement of the left or right hand or shoulder. This

instruction was extinguished during a delay period. A generic “Go” cue, visually identical across trial conditions, prompted movement. From initial pilot data we knew that all hand and shoulder movements evoked activity in the population; however, we were unsure how the different conditions mapped onto individual neurons. Shoulder and hand movements are frequently performed together opening the possibility that hand and shoulder movements would frequently be localized to the same neural population. We therefore introduced speech as a fundamentally different action that could provide an additional movement for comparison. During the speech conditions, N.S. simply said “left” or “right” as instructed. Eight repetitions of each trial type were pseudorandomly inter-leaved such that one repetition of each condition was performed before repeating a condition.

Figure 3.1B-E show several well-tuned example units that highlight how neurons commonly coded for a complex assortment of different condition types. For instance, Example B codes for movements of the right hand, whether or not the movement was imagined or attempted. Example C codes exclusively for attempted movements of the left hand. Example D responds similarly for imagined actions of the left or right hand, but not attempted actions. Example E codes for when N.S. spoke “left.”

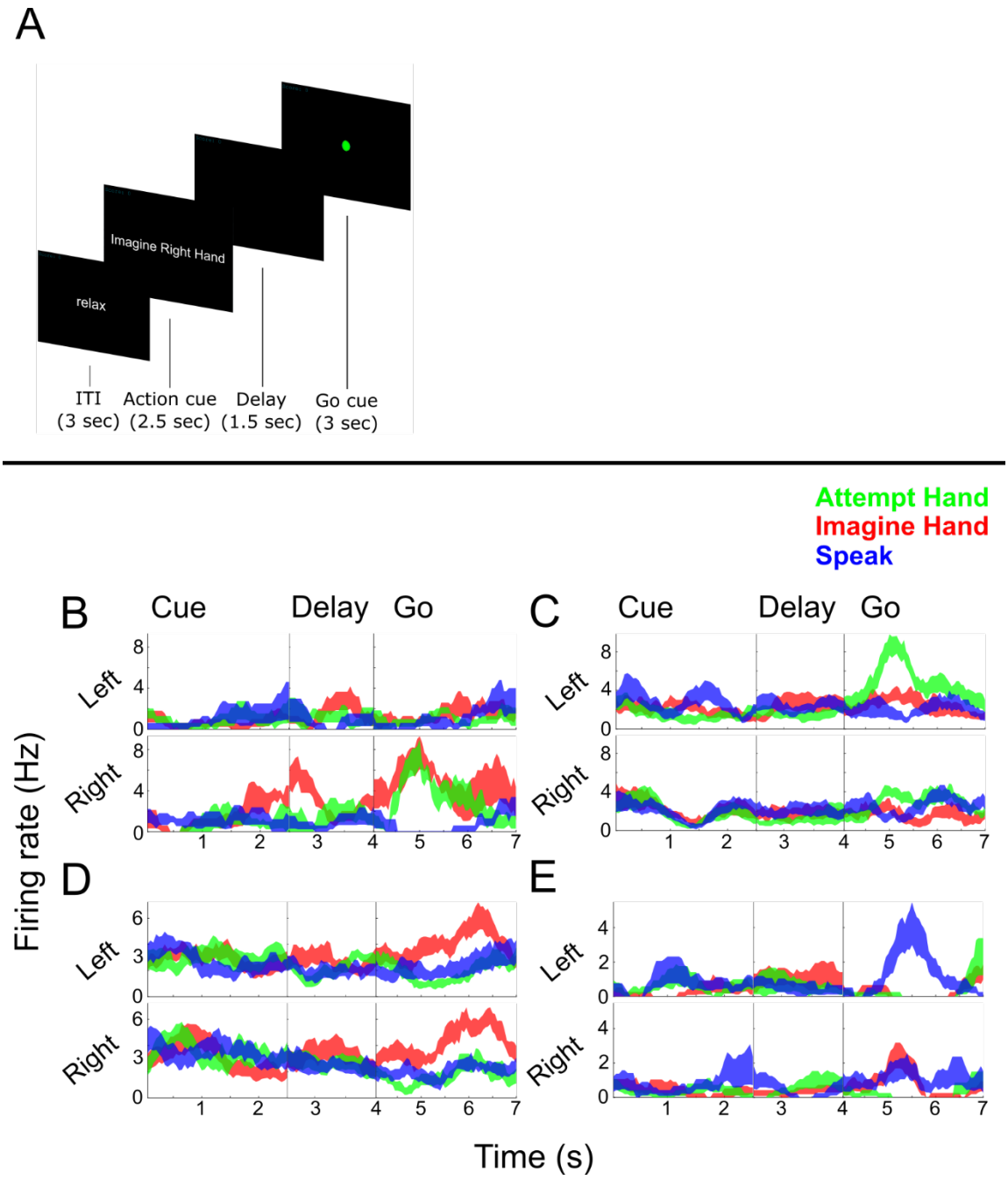

Figure 3.1. Neurons in PPC Exhibit Mixed Selectivity to Movement Variables

(A) Delayed movement paradigm. N.S. was cued as to what kind of movement to perform (e.g., imagine/attempt left/right hand/shoulder) and then cued to perform the movement after a brief delay. See Methods for more details. (B-E) Single unit example responses over time (mean ± SEM) demonstrating diverse coding to the different conditions.

.

To better understand the strength of tuning in the population to each condition, we fit a linear model to each neuron that explained firing rate relative to baseline (taken as the firing rate during the ITI) as a function of each task condition for both the Go and Delay phases. All 8 primary movement conditions as well as speaking were represented in the neural population (Figure 3.2A). We also examined the magnitude of information content for recorded units by computing the area under the receiver operating characteristic curve (AUC) generated when comparing the Go/Delay period activity to ITI activity for each condition separately (Figure 3.2B). While there were significant differences between specific pairwise comparisons (Figure S 3), results across these measures were comparable overall.

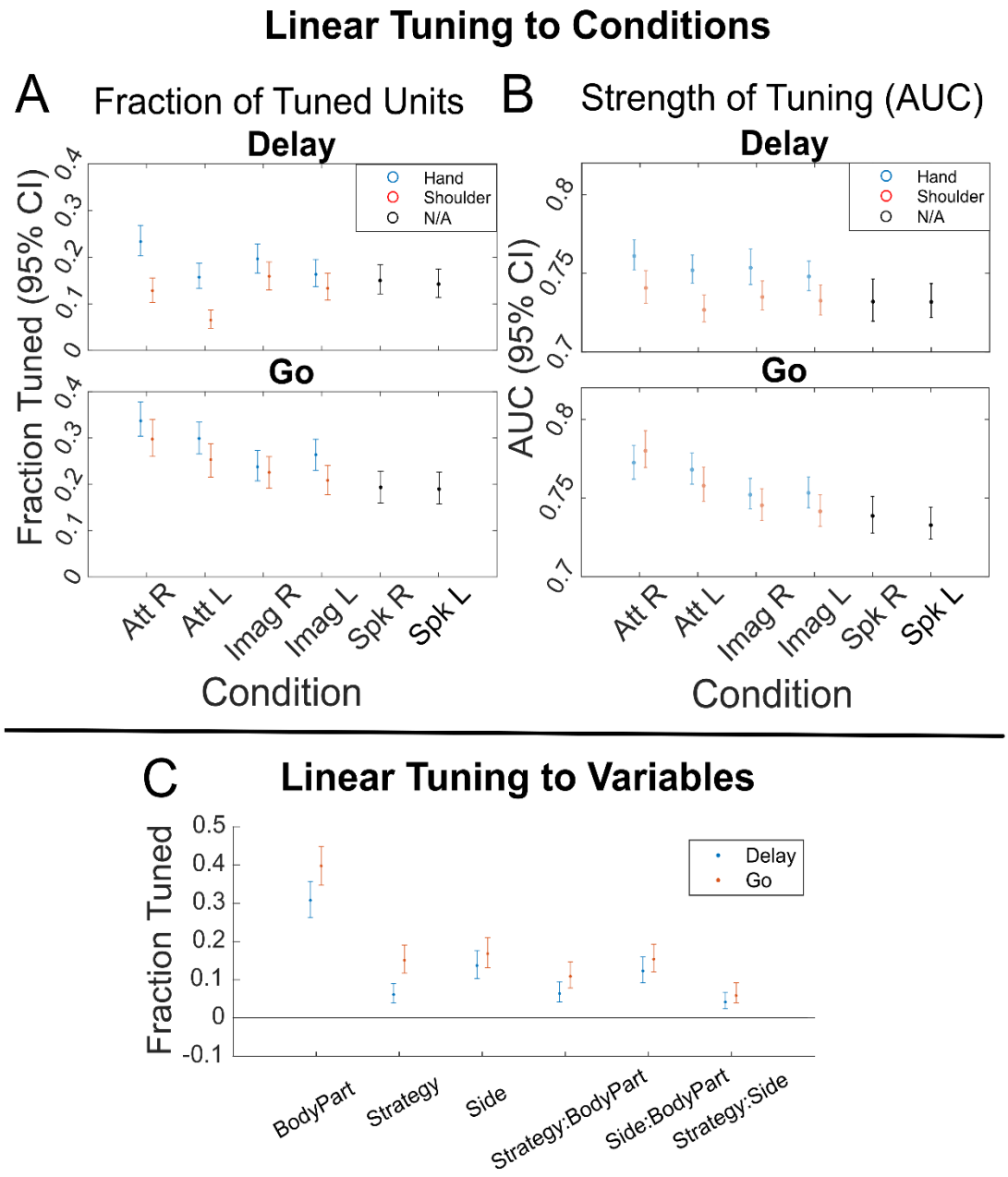

Figure 3.2. Significant Tuning to Each Movement Condition

(A) The fraction of units in the population tuned for each condition in the Delay and Go phases, separated by body part and body side (95% confidence interval). A unit was considered tuned to a condition if the beta value of the linear fit for the condition (Linear Analyses 1 Methods) was statistically significant ($p < 0.05$, uncorrected). See also Figure S 3 for pairwise comparisons between conditions and Figure S 5 for results of individual sessions. (B) The magnitudes of the units' tuning to each condition in the Delay and Go phases, as defined by the area under the receiver operating characteristic curve (AUC) between Delay/Go and ITI activity, separated by body parts (95% confidence interval). Only significant AUC values were included in analyses (shuffle test, $p < 0.05$ uncorrected). See also Figure S 4 for the AUC values of excitatory (positively tuned) and inhibitory (negatively tuned) units presented separately, as well as Figure S 3 for pairwise comparisons between conditions. (Att R = Attempt Right, Att L = Attempt Left, Imag R = Imagine Right, Imag L = Imagine Left, Spk R = Speak Right, Spk L = Speak Left). (C) Fraction of units with significant tuning to each motor variable and the interaction terms for both the Delay (blue) and Go (red) phases, as opposed to the 8 movement conditions in (A) ($p < 0.05$, uncorrected 95% confidence intervals, see also Linear analysis 2 in Methods).

We performed a second type of linear analysis, fitting a linear model that explained the firing rate relative to baseline as a function of the three motor variables of strategy, body side, and body part (Figure 3.2C). This analysis revealed an asymmetry in how body part is represented compared to body side or strategy. In particular, relatively larger non-linear interactions of strategy and body side with body part indicate that body part may in some way structure the functional responses to the other variables, a point we directly address below.

We found significant differences in mean firing rates between hand and shoulder movements (hand greater than shoulder, MANOVA $p = 5.614e-8$) and attempted and imagined movements (attempt greater than imagine, $p = 0.0020$), and no significant differences in mean firing rates between left- and right-sided movements ($p = 0.2951$). Comparing the firing rates of all 8 movement conditions (pooled together) with the firing rates of the speech conditions, we found a significant bias towards hand and shoulder movements over speaking (t-test $p = 4.7048e-8$).

How are these different motor representations coded with respect to each other in the same region of cortex? Figure 3.3 shows five possibilities: (1) the 8 movement condition representations could be anatomically segregated from each other, with a highly specialized sub-population of neurons dedicated to each (sparse mixed selectivity, Figure 3.3A); (2) an organization similar to one, save that some variables are subordinate to others. For instance, imagined movements may be a subset or suppressed version of attempted movements (Figure 3.3B); (3) highly specialized sub-populations are tuned to each motor variable class exclusively (body part, body side, strategy) (pure selectivity, Figure 3.3C); (4) each motor variable class

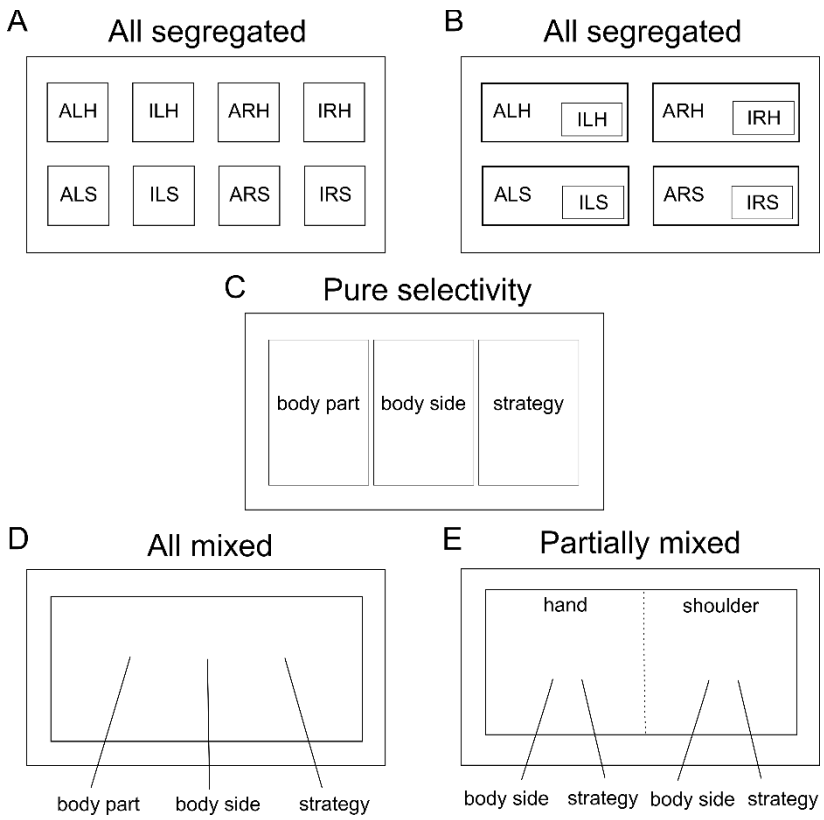

Figure 3.3. Possible Organizational Models of Neural Representations

(A) The neurons coding for each condition are anatomically segregated, i.e., distinct, non-overlapping networks. (ALH = Attempt Left Hand, ILH = Imagine Left Hand, ARH = Attempt Right Hand, IRH = Imagine Right Hand, ALS = Attempt Left Shoulder, ILS = Imagine Left Shoulder, ARS = Attempt Right Shoulder, IRS = Imagine Right Shoulder). (B) Conditions can be overlapping such that the responses to some conditions are subsets or weak versions of others, e.g., imagined movements being subsets of attempted movements. (C) Neurons coding each of the motor variables (body part, body side, and strategy) are anatomically segregated. (D) The neural population exhibits mixed selectivity, with individual neurons showing tuning to various conjunctions of variables. (E) The neural population exhibits partially mixed selectivity, with the mixing of representations being dependent on the variables under investigation. Here, hand and shoulder are mixed leading to orthogonal coding of effectors (functional segregation), however, the other variables (body side and strategy) are mixed only within, but not between, effectors. This model is consistent with the results observed in this study. Note that solid lines in this diagram indicate anatomical boundaries of neural populations while dotted lines indicate functional boundaries/segregation.

could be randomly mixed together (Churchland and Cunningham, 2015; Fusi et al., 2016) (all mixed, Figure 3.3D); (5) some variables may be randomly mixed while others are organized with more structure (partially mixed, Figure 3.3E).

We first performed a degree of specificity analysis (Figure 3.4) to determine the following: (1), whether highly specialized sub-populations of neurons are dedicated to each movement type, and, (2), whether some variables exist as subsets or suppressed versions of other variables. A

specificity index was computed as the normalized difference in beta values between motor variables for each neuron (taken from the linear models described above). Values near zero indicate equivalent neural responses to the two conditions being compared while values near 1 (or -1) indicate exclusive neural responses for one condition. By proposition one, we would expect values to be clustered near 1 (or -1) as, e.g., either a neuron is tuned to the right side or the left side. By proposition two, we would expect strong biases such that values would be clustered on one side of the range (between 0 and 1 *or* 0 and -1) as, e.g., a neuron tuned to imagine movement should be better (or equivalently) tuned to attempted movement. Inconsistent with these proposals, we found that specificity values were distributed over the full range (Figure 3.4A-F). For instance, despite a small population bias for attempted movements, a sizable proportion of neurons were exclusively, or more strongly activated for imagined movements (Figure 3.4AB; see Figure 3.1D). The neural representation of motor imagery is thus not a subset, or less strongly represented version, of motor execution. Likewise, many neurons showed preferential coding for the left hand (Figure 3.4CD) even with a population bias for the right hand. There was a strong specificity bias towards the movement conditions (imagined or attempted movements of the hand or shoulder) over the speech conditions (Figure 3.4GH). This is expected given that speech tuning is found in a smaller proportion of neurons in a weaker fashion (Figure 3.2). Of special note, the results here are very similar for both movements of the shoulder (above the level of injury) and movements of the hand (below the level of injury). Thus, movements below and above the level of injury are coded in a similar manner.

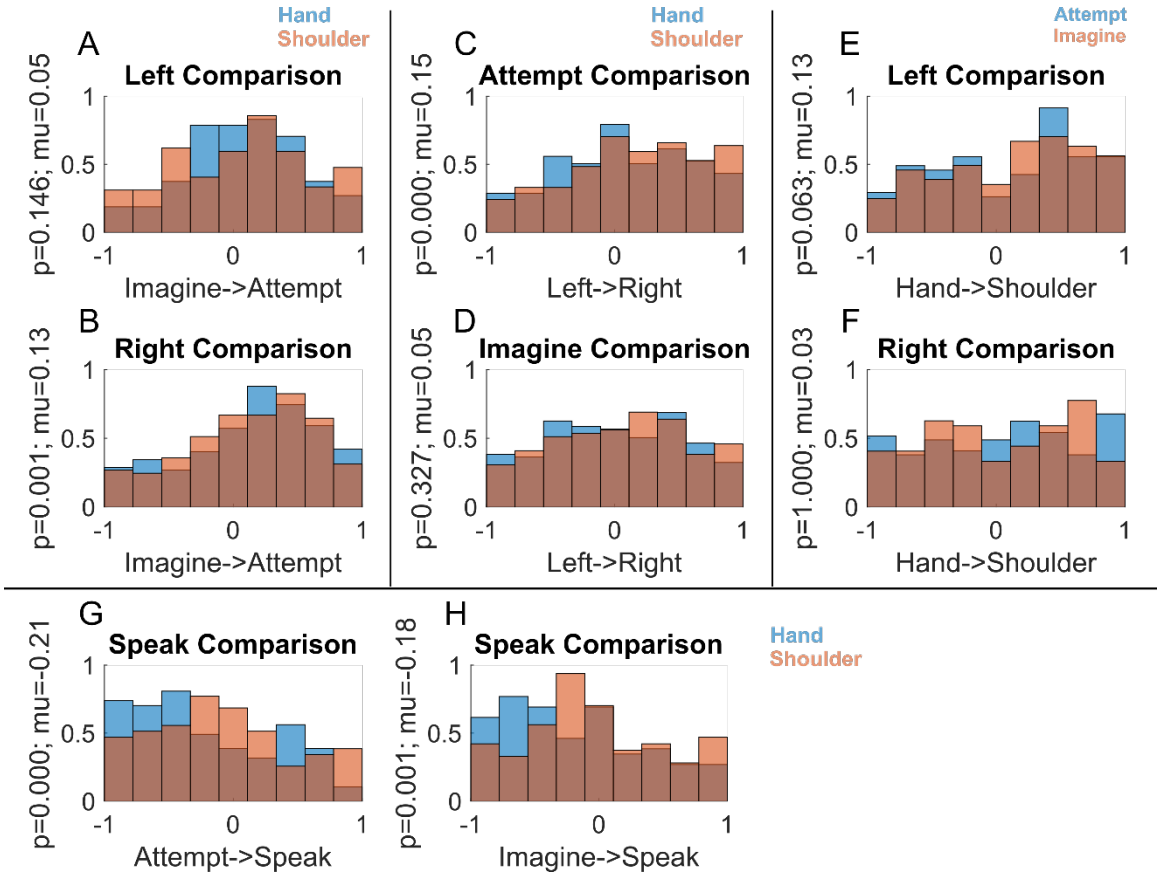

Figure 3.4. Specificity of Coding for Motor Variables

Each panel (A-F) shows the degree to which neurons code one variable exclusively, its opposite, or respond similarly for both. Only units with significant modulation for at least one condition in the comparison are included in the analyses ($p < 0.05$, Bonferroni corrected). (A-B) Distribution of the degree of specificity to the imagine or attempt strategies in the population during trials using different sides, showing only units responsive to one or both strategies. (C-D) Distribution of the degree of specificity to the left or right side in the population for different strategies. (E-F) Distribution of the degree of specificity to the hand or shoulder in the population during trials using different sides. (G-H) Distribution of the degree of specificity to attempted/imagined movements compared to speaking.

We failed to find complete specialization of function across the population for single units, and the distributed and overlapping nature of responses makes it difficult to find structure in the responses of individual neurons. We therefore turned to population-based analyses to more readily identify how the different conditions are encoded with respect to each other. We measured all pairwise correlations between population responses for each of the eight movement conditions and looked for systematic structure in how the different motor variables (body part, body side, cognitive strategy) were coded (Figure 3.5). Correlation was used as a measure of similarity over other distance measures such as Euclidean or Mahalanobis distance

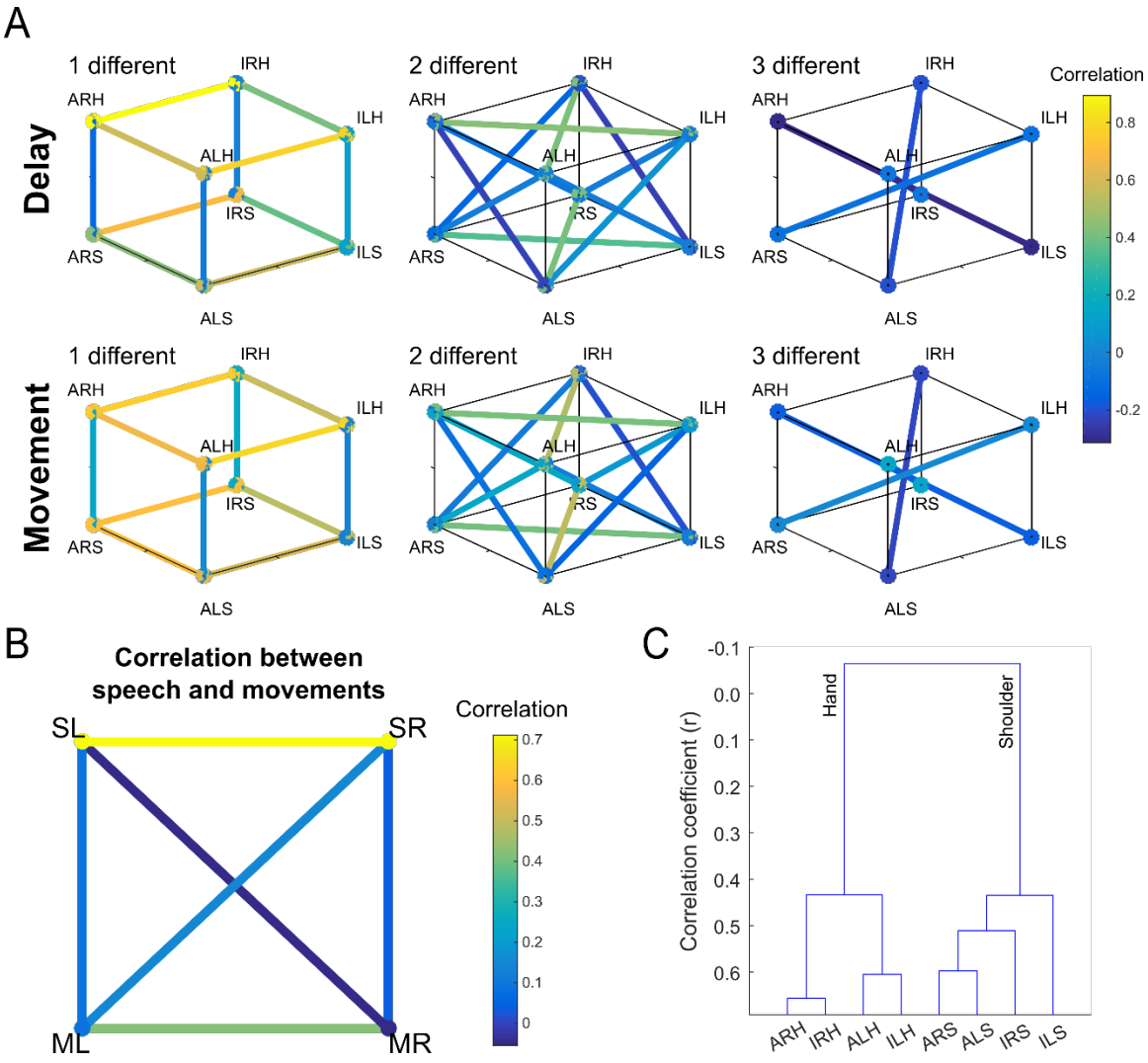

Figure 3.5. Functional Relationships between Movement Conditions

(A) Similarity between population level neural responses for each movement condition. Pairwise comparisons are separated by the number of motor dimensions that differ in the comparison (left to right) and task phase (movement or delay). Similarity measured as the pairwise correlation between movement conditions. (ALH = Attempt Left Hand, ILH = Imagine Left Hand, ARH = Attempt Right Hand, IRH = Imagine Right Hand, ALS = Attempt Left Shoulder, ILS = Imagine Left Shoulder, ARS = Attempt Right Shoulder, IRS = Imagine Right Shoulder). See also Figure S 5 for results of individual sessions. (B) Correlations between four movement types: left and right movements (averaged across both strategies), and speech. (SL = Speak Left, SR = Speak Right, ML = Movement Left, MR = Movement Right). (C) Dendrogram summarizing the structure apparent in A, namely strong segregation by effector.

because the sign of the correlation is potentially informative of the underlying structure. For example, two conditions represented by distinct neural populations (e.g., sparse mixed selectivity) would manifest as a negative correlation between the two conditions, while a positive correlation would indicate a degree of overlap between the populations. Asymmetric relationships between the different variables were immediately apparent.

Correlations between conditions that differed in body side or cognitive strategy were high if the comparisons were made within a body part. In stark contrast, correlations between conditions that differed in body part were low even if cognitive strategy and body side were held constant (Figure 3.5A and Figure 3.6A). Low correlation between body parts was also apparent when comparing speech with shoulder or hand (Figure 3.5B). Such low correlations despite activating overlapping neural populations are a signature of network responses that occupy distinct neural subspaces thus minimizing crosstalk during planning and execution epochs (Churchland and Cunningham, 2015; Kaufman et al., 2014). Here the same principal may be at play for cortical representations of different effectors in an overlapping neural population. We term this "functional segregation" of body parts. That the functional organization is based around effector is especially apparent when the distances between conditions were hierarchically clustered (Figure 3.5C), with body part being the primary differentiating variable. Further, for a given body part, movements with more shared traits are coded more similarly than movements with fewer shared traits (Figure 3.6B). For instance, a neuron tuned to *imagined* left hand movements was more likely tuned to *imagined* right hand movements (but not *attempted* right hand movements). Likewise, a neuron tuned to *right* hand imagined movements was likely to be tuned to *right* hand attempted movements (but not *left* hand attempted movements). This functional segregation likely accounts for the non-linear interaction terms of Figure 3.2C.

Neural differences between hand and shoulder movements may be driven by the fact that the hand is below the level of injury while the shoulder is above the level of injury: in this case, proprioceptive feedback or long-term effects from the injury might be the primary difference.

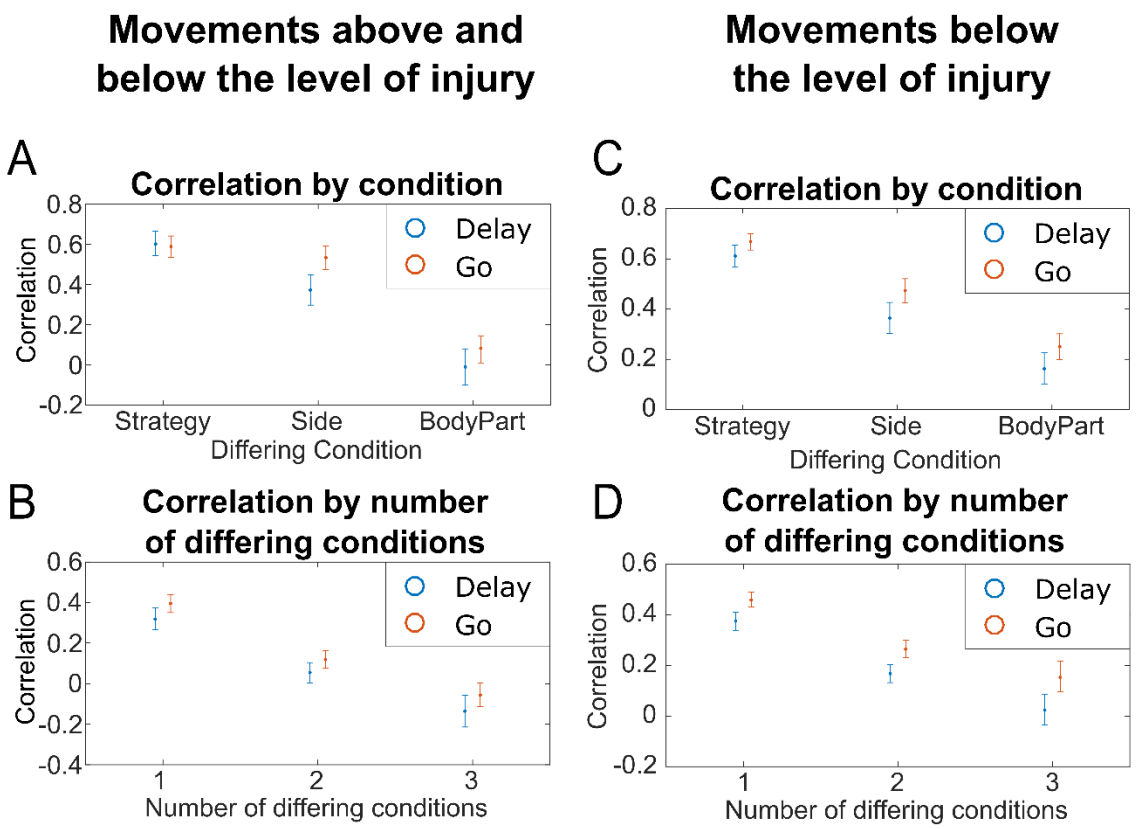

Figure 3.6. Segregation by Body Part

(A) Average correlation between movement conditions differing by exactly one task variable and grouped by the differing condition (e.g., for strategy, the average correlation of all movement condition pairs differing only by strategy). Intervals represent the 95% confidence intervals. (B) For movements above and below the level of injury, average correlation between movement conditions in the Delay and Go phases grouped by the number of differing traits (average of each cube in the movement phase). Intervals represent the 95% confidence intervals in the correlations. (C-D) Same as (A-B) but with shoulder shrug movements replaced with shoulder abduction movements (a movement below the level of injury).

To address this issue, we replaced shoulder shrugging movements with shoulder abduction movements (shoulder abduction resulted in no overt movement) and repeated the correlation analyses. The results are similar when both body parts are chosen to be below the level of injury (Figure 3.6CD). In particular, the largest degree of separation exists between body parts.

Functional segregation of body parts should lead to minimal shared information about other motor variables when compared across body parts. The motor dimensions can be thought of as

categorical variables with two levels (e.g., body part has the levels of shoulder and hand). Given functional segregation, a classifier trained on one level ("Level A") should fail to generalize to the other category ("Level B") and vice versa (Figure 3.7A). Alternatively, for highly overlapping representations, a classifier trained on Level A should generalize to Level B and vice versa (Figure 3.7B). For example, given functional segregation between hand and shoulder, the neural signature that differentiates right from left sided movements for the hand should fail to generalize to the shoulder. We tested for this possibility by looking at patterns of generalization across trained classifiers. The results of such an analysis are shown in Figure 3.7C-H. For Figure 3.7C, we trained a linear discriminant classifier on all shoulder movement trials to differentiate between left and right-sided movements, regardless of strategy. The decoder performed well within its own training data as expected (leave-one-out cross-validation, Figure 3.7C, left blue bar), but performed at chance differentiating left from right-sided movements for hand trials (Figure 3.7C, right blue bar). The reverse was true when applying a classifier trained on hand trials to shoulder trials (Figure 3.7C, orange bars). Likewise, Figure 3.7D shows that a decoder trained to differentiate strategy using shoulder trials failed to generalize to hand trials, and vice versa. In contrast, decoders trained to differentiate strategy or body part *were* able to generalize and perform well across different body sides (Figure 3.7EF) and different strategies (Figure 3.7GH). Body part differences exhibit functional segregation while cognitive strategy and body side do not.

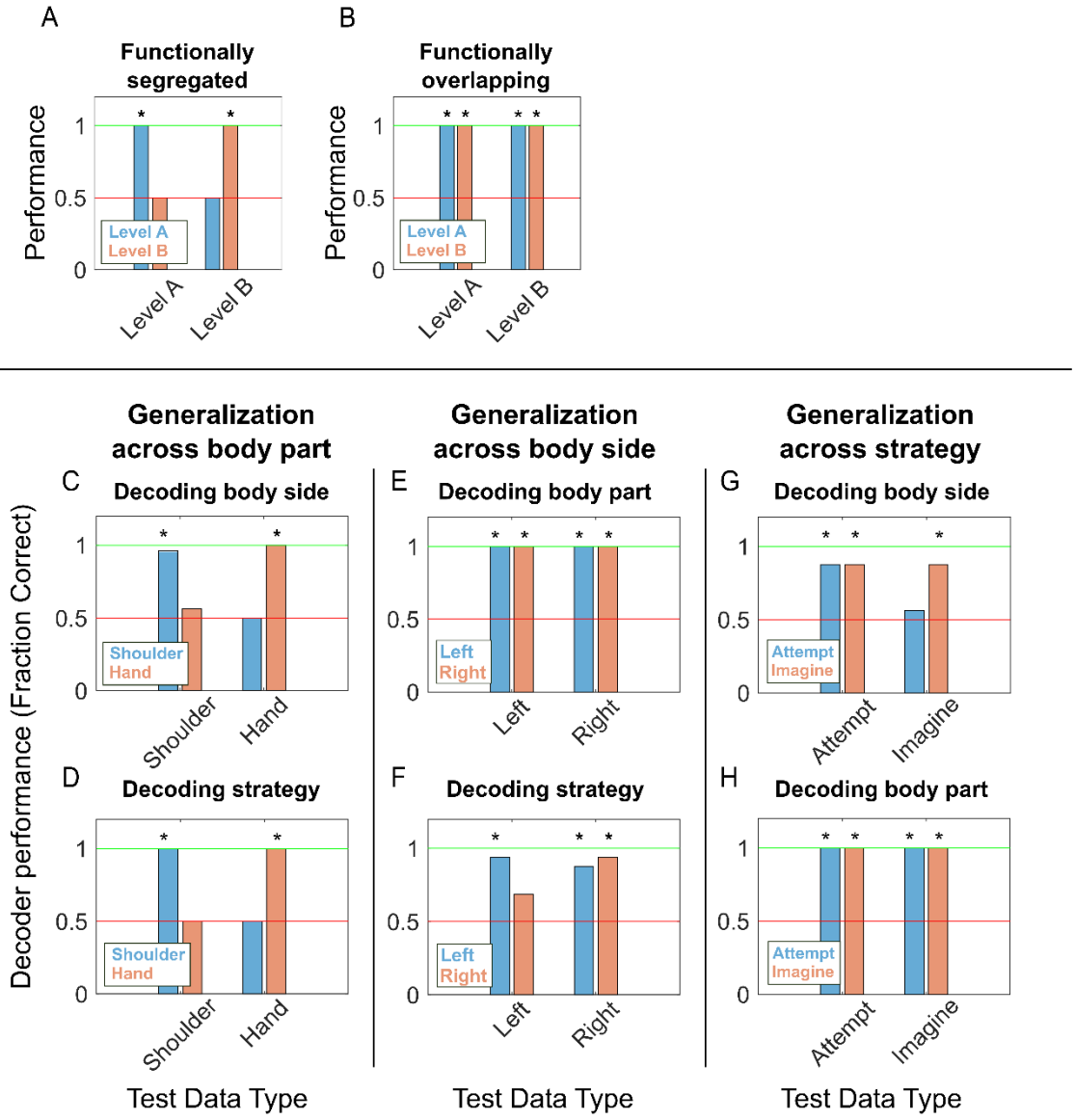

Figure 3.7. Representations of Variables Generalize across Side and Strategy, not Body Part

Schematic illustrating expected classifier behavior if variables are functionally segregated (A) versus overlapping (B). (A) Functional segregation within a variable (e.g., body part) implies that a classifier trained to differentiate the levels of one dimension (e.g., right from left) will not generalize across the levels of the dimension of interest (e.g., from shoulder to hand) resulting in chance performance. (B) In contrast, functional overlap implies generalization resulting in above chance performance when comparing classifier performance across levels. (C) Performance of decoders trained on data split by body part for classifying the body side. Blue/orange bars represent the performance of the decoder trained on shoulder/hand movement data. Horizontal axis labels represent which body part's data each decoder was tested on. Performance was measured as the fraction of trials accurately classified by the decoder, with in-sample performance determined by cross-validation. Asterisks represent performance significantly different from chance, as determined by a rank shuffle test. The red line represents chance performance level (0.5) while the green line represents perfect performance (1.0). (D) Similar to (C) but decoding strategy instead of body side. (E-F) Similar to (C) but with data split by body side and decoding for body part and strategy, respectively. (G-H) Similar to (C) but with data split by strategy and decoding for body side and body part, respectively.

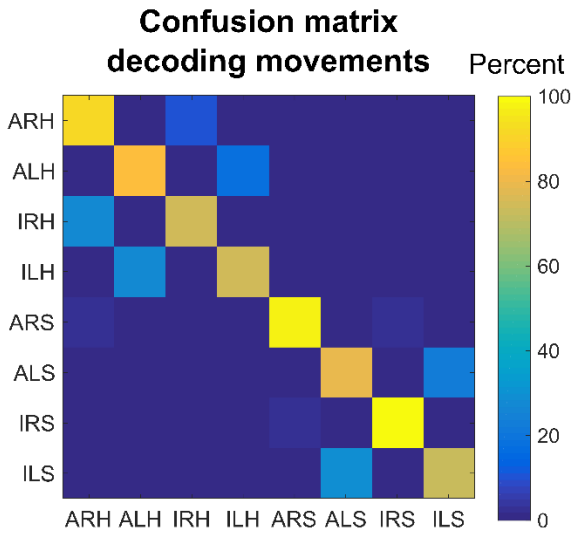

Figure 3.8. All Movement Variables Decodable from the Population

Confusion matrix for cross-validated classification of the eight movement conditions. (ALH = Attempt Left Hand, ILH = Imagine Left Hand, ARH = Attempt Right Hand, IRH = Imagine Right Hand, ALS = Attempt Left Shoulder, ILS = Imagine Left Shoulder, ARS = Attempt Right Shoulder, IRS = Imagine Right Shoulder).

Given that some motor variables are similar in their neural encoding, is it possible to decode the body part, body side, and cognitive motor strategy from the neural population? We constructed a neural classifier to differentiate all conditions (Figure 3.8). Cross-validated classification performance was high. However, as expected, misclassification tended to occur between conditions with more variables in common. This is especially true between attempted and imagined movements as predictable from the high degree of similarity in the neural responses (Figure 3.6A).

3.4 Discussion

We tested how a variety of motor variables were coded at the level of single neurons in human AIP. This allowed us to address several questions about how intent is coded in human AIP and to better understand how the motor variables are coded with respect to each other.

3.4.1 Effector specificity in PPC.

Classically, the regions around the IPS have been viewed as organizing around the control of different effectors such as the eye, hand, and arm. In a recent challenge to the centrality of an effector-based organization, Medendorp and colleagues have found that effector-specificity in the BOLD response of fMRI is much more pronounced between the hand and eye than the hand and other body parts arguing that effectors as such are not differentiated in the planning regions of PPC (Heed et al., 2011a). In line with these results, we found significant numbers of neurons tuned to movements of the hand and shoulder in a small patch of AIP. However, unlike the response at the level of voxels, the neural response to each effector was functionally segregated. Thus, while our results challenge the idea of strict anatomical segregation of effector representations across cortical areas, we do find local functional segregation of effectors within a cortical field. The current findings suggest that effector-specificity at the global anatomical scale could be thought of in terms of relative emphasis rather than strict specialization in humans.

In non-human primates (NHPs), a global organization for eye and arm movements is supported by greater planning activity of single neurons for reaches in the parietal reach region (PRR) and saccades in the lateral intraparietal region (Cui and Andersen, 2007; Hwang et al., 2012; Quiroga et al., 2006; Snyder et al., 1997; Snyder et al., 1998). Reversible inactivation of PRR produces reach specific deficits and LIP a bias toward saccade deficits (Christopoulos et al., 2015; Kubanek and Snyder, 2015; Yttri et al., 2014). A grasp specific deficit has been reported for AIP (Gallese et al., 1994). These results indicate that, at a global level, there is functional specificity by effector in non-human primates, and fMRI studies suggest a similar global specialization. However, these

areas also communicate with one another. For instance, inactivation produces a reach deficit in PRR when reaches are made alone, but both reach and saccade deficits when combined hand-eye movements are made (Hwang et al., 2014). Thus, the degree of effector overlap in AIP in human may reflect the coordination of movement and communication between effector specific areas.

An advantage of our human study is that the participant can perform a large number of tasks by verbal instruction. In NHP studies, the animals must be trained for long periods and thus the number of tasks and task variables are generally limited per study. Interestingly, area LIP has been studied by a number of groups using a number of different tasks. As a result, LIP has been found to modulate activity for tasks examining movement planning, attention, categorization, and decision making resulting in a variety of proposals for its function (Andersen and Cui, 2009). It may be that the large number of variables to which human AIP is selective may be a reflection of the versatility of using different tasks and that both human and NHP PPC areas are modulated by a very large number of variables. Indeed, several NHP studies in AIP have reported overlapping populations of cells tuned to grasp type and reach target consistent with mixed selectivity between effectors as presented here (Asher et al., 2007; Fattori et al., 2009; Lehmann and Scherberger, 2013, 2015).

Differences between effector segregation in human and NHP studies of PPC may be a result of possible lack of homologies between human AIP of the current study and AIP of NHPs. In fact, we do not know the extent or number of grasp related areas defined by single neuron recordings in human IPS and whether there are grasp regions in humans that do not exist in

NHPs. Finally, the lack of strict anatomical segregation of effectors may point towards a global topographic organization governed around more behaviorally meaningful aspects of behavior such as manipulation, reaching, climbing, and defense (Graziano and Aflalo, 2007; Jastorff et al., 2010). Whichever possibilities outlined above account for the large number of variable encodings in human AIP, an exciting aspect of our results is that they open the possibility of decoding movements of many body parts from one small patch of cortex.

3.4.2 Asymmetric coding of motor variables and functional segregation of body parts

Recently there has been increased interest in not only the types of variables that are coded in a cortical region, but also how these variables are coded with respect to each other in an effort to understand the underlying logic of the computations performed within a cortical field (Fusi et al., 2016; Raposo et al., 2014). For instance, several papers have shown that higher cortical areas like PPC and prefrontal cortex may employ a computational strategy by which response variables are randomly mixed (Raposo et al., 2014; Rigotti et al., 2013). While such a coding scheme can give rise to complex and difficult to interpret representations at the level of single neurons, the population code is information rich and enables simple linear classifiers to decode any variable of interest. In these papers, it was shown that response variables were randomly distributed across neurons, as illustrated in Figure 3.3D. Our data provides insights into understanding population coding by demonstrating that in human AIP certain response features can be seemingly randomly distributed across the population while others are not. In particular, we find that coding for body part is uncorrelated in the sense that across the population, knowing that a neuron is tuned to shoulder movements provides little to no information about whether the neuron is tuned to hand movements (or speech; Figure 3.4). This is true even if you know other attributes of the movement, such as whether the movement was imagined or

attempted or performed with the right or left side of the body. In contrast, when comparing within the same body part, knowing a neuron is tuned to movements of the right side makes it highly likely that the neuron will be tuned to the left side as well. The same is true for imagined and attempted movements. Thus while some variables seem to be randomly distributed across the population (e.g., body part) the relationship between other variables (e.g., body side, mental strategy) is organized in relationship to a third variable (body part). This effectively leads to functional segregation of body part at a population level. Such functional segregation between body parts is very similar in principal to the relationship between planning and execution related activity that has recently been described in frontal motor areas (Churchland et al., 2010; Kaufman et al., 2014) where planning activity fails to excite subspaces that are hypothesized to produce muscle output.

But why are some variables functionally overlapping while others are functionally segregated? One possible answer is computational savings. Overlapping activity at the level of the population may be rooted in shared computational resources. For example, many computations related to planning and executing grasps including object affordance processing as well as basic kinematic processing would be similar for the right and left hand. Motor imagery has also been hypothesized to engage internal models used for sensory estimation during overt execution (see below) and thus imagery and execution should rely on largely overlapping computations. Thus despite the potential computational benefits to random mixing of variables (Fusi et al., 2016), the computational savings of overlapping resources for certain classes of computations may outweigh losses in the total information the population encodes.

Another possibility is that the highly overlapping representations provides part of the neural substrate through which transfer of learning occurs. Motor skills learned with one hand frequently result in improvements in performance with the other hand (Amemiya et al., 2010). Likewise, use of motor imagery is found to improve performance during motor execution (Dickstein, 2007). One possibility is that overlapping networks would be able to facilitate this sort of transfer of learning. For example, repeatedly imagining a movement with the right hand would recruit a similar network as executing a movement with the right hand, making any neural adaptation from learning the movement more likely to transfer between the strategies.

Despite the greater functional overlap between body side and strategy, it is important to note that all the tested movement conditions are still differentiable from each other (Figure 8). Interestingly, this greater overlap for body side may explain why patients with motor deficits often "mirror" movements in a contralateral limb. In cerebral palsy, for example, patients making a grasp with their left hand often mirror the movements with their right hand (Kuhtz-Buschbeck et al., 2000).

A point of note is that the movements selected in this study (hand squeezes and shoulder shrugs) are not necessarily the best exemplars of movements of the respective body parts. Different combinations of hand or shoulder movements may have slightly more or less overlap. Bilateral symmetric hand squeezes and shoulder shrugs occur more naturally than "squeeze-and-shrug" actions, action not part of the natural movement repertoire. The statistical frequency with which different body parts are moved together could also affect the degree of functional overlap between the body parts. A better understanding of how different exemplars

of movements across different effectors relate will be important in understanding the functional organization of motor actions in AIP.

3.4.3 Attempted and imagined movements in human AIP after long-term injury

In this study, we looked at neural coding of imagined and attempted actions above and below the level of injury in a paralyzed individual. By current theory, imagined movements may represent the simulation of an internal model of the arm, a model that also forms the basis for sensory estimation during overt forms of behavior (Gail and Andersen, 2006; Jeannerod, 1995; Mulliken et al., 2008b). In broad strokes, this theory predicts that neural representations of imagined and overt movements should have a high degree of similarity given the shared neural substrate, but also be different given the absence of movement during imagery (Jeannerod, 1995; Munzert et al., 2009). Our results support this view insofar as we demonstrate the high degree of functional overlap between imagined and attempted movements. However, we also show neural differences between imagined and executed movements persist and are highly similar even after long-term injury and disuse (see Figure 3.4 and Figure 3.5). Such a preserved distinction does not immediately follow from the proposal that the primary difference between imagined and executed movements is the actual movement itself. However, the observation that neural coding differences persist even when injury results in a lack of overt movement during attempted actions is inconsistent with the proposal that the primary difference between imagined and executed movements is the actual movement itself (Jeannerod, 1995). Further, the patterns of similarities and differences in how the population codes mental strategy and body side—for instance, the preference for attempted over imagined movements for the right but not left side of the body (Figure 3.4A versus Figure 3.4B) —demonstrate that higher-order population structure is conserved following injury. This suggests that preservation of motor

intention signals enables successful BMI functionality many years post-injury (Aflalo et al., 2015). A better understanding of how different cortical subregions maintain representations of motor intent post-injury may help inform choice of implant sites as a function of time post injury.

These results demonstrate for the first time that networks activated during attempted actions are highly overlapping with networks activated during imagined actions at the level of populations of individual neurons, and that the correspondence between actions is body part specific (hand and shoulder). However, there is a symmetry in our results such that networks activated during right hand actions are highly overlapping with networks activated for left hand actions, and the correspondence between right and left actions are strategy specific (e.g., right-side actions look more like left-side actions using the same strategy). In other words, the relationship between imagined and attempted actions is similar in basic form to the relationship between left and right sided actions although the degree of overlap is greater for strategy.

The current experiment was performed in the larger context of a brain-machine interface clinical (BMI) trial. We have previously shown that a paralyzed patient can use motor imagery to control a robotic limb (Aflalo et al., 2015). Is the use of motor imagery the best method for the user of a BMI to modulate their own neural activity? Alternatively, it is possible that attempted movements somehow better engage or otherwise enable the user to control an external device. Here we show that the distinction between imagined and attempted actions is preserved, even in limbs for which no movement is possible. Future work is needed to determine whether these differences translate into performance differences during closed-loop neural control.

3.4.4 Orofacial coding in human AIP

We included speech conditions in which N.S. spoke “left” and “right” as a third fundamentally different movement. A smaller proportion of neurons were tuned more weakly to speech acts, demonstrating that not all actions are coded in an equivalent manner in AIP (Figure 3.2 and Figure 3.4). This task was not designed to understand the functional significance of “speech” tuned units, but one possibility is that these neurons code for orofacial movements and may form the building blocks for more complex coordinated movements of behavioral relevance such as coordinated movement of the hand to the mouth for feeding or tearing open a bag of chips with your mouth. It is also possible that such orofacial tuning coordinates “grasping” actions performed with the mouth (Jastorff et al., 2010).

4 Mixed representations in closed-loop cortical control

4.1 Introduction

An important finding in systems neuroscience is that cortical neurons exhibit mixed selectivity, i.e., where individual neurons are tuned to multiple variables in idiosyncratic ways. Early studies examining the representation of extrapersonal visual space in the posterior parietal cortex (PPC) of non-human primates (NHPs) showed that individual neurons, instead of having eye position invariant receptive fields, combined retinotopic receptive fields with eye position signals (Andersen and Mountcastle, 1983). These two signals often interacted multiplicatively and were referred to as "gain fields" (Andersen et al., 1985). Neural networks trained to transform retinotopic receptive fields to craniotopic receptive fields also formed gain fields similar to the neural data (Zipser and Andersen, 1988). Moreover, PPC neurons also mixed head position signals (Brotchie et al., 1995) and vestibular signals (Snyder et al., 1998) for potential transformations to body and world coordinates. Recent studies have found random mixing of auditory and visual signals in rat PPC from retinotopic receptive field and eye position signals (Raposo et al., 2014) as well as random mixing of sensory and task related signals in NHP prefrontal cortex (PFC) (Rigotti et al., 2013). Mixed selectivity allows a relatively small population of neurons to encode a large variety of variables (Fusi et al., 2016). Consistent with those results, our lab has also recently found mixed selectivity, with multiple body parts (effectors) and cognitive strategies represented within a small patch of human PPC (Zhang et al., 2017). In particular, imagined and attempted movements of the left and right hands and shoulders were all decodable and differentiable from each other in the human anterior intraparietal area (AIP).

The ability to decode these mixed variables is particularly promising for brain-machine interfaces (BMIs) and suggests the possibility of decoding many variables using a single array. BMIs allow a user to control external prosthetic devices by directly decoding motor intentions from patterns of neural activity learned during an initial "training" dataset. Previous studies on mixed selectivity, however, have studied neurons in the absence of any BMI online control, only examining the data that would be used to train a BMI decoder. Thus, there remains an open question: how do the mixed representations of different variables in AIP compare between training and online control? More specifically, is the *structure* of the representations maintained?

The answer is not immediately obvious. Despite studies showing that the tuning of specific neurons can change from training to online control (Chase et al., 2009; Cunningham et al., 2011), BMI control based on representations learned in open-loop/training is still possible. In a recent study we found that motor imagery of arm movements decoded from PPC can be used by a subject to perform closed-loop cortical control of a computer cursor or robotic limb (Aflalo et al., 2015). Other studies in non-human primates (NHP) have demonstrated BMI control via effectors in other parts of PPC as well (Graf and Andersen, 2014; Revechkis et al., 2014). These studies, however, focused on BMI control with respect to a single effector or with different effectors in different sessions, and did not directly compare the representations of different effectors between training and online control. It is possible that the structure of the representations is maintained between training and online control. Alternatively, it is also possible that during online control the different representations collapse into an effector-independent intention, attention, or goal signals.

In this study, we focus on two questions: (1) how is the structure of the movement representations maintained between training and online control, and (2) how does the degree of maintenance influence the control performance of the different types of movements?

To answer these questions, we test the BMI control performance of a C3/C4 tetraplegic participant in a 1D cursor control task, decoding from single unit activity recorded from a small 4x4 mm patch of AIP. The participant controlled the cursor using imagined or attempted movements of the left (ipsilateral) or right (contralateral) hand. We find that all four of the tested movement conditions remain differentially represented during online control, with the structure of the representations largely maintained. While attempted right (contralateral to the implanted hemisphere) hand movements performed significantly better than the other movements, the effect was primarily driven by more units being tuned to the condition rather than differences in how well the representations were maintained.

4.2 Methods

4.2.1 Subject Details

Subject NS is a 59 year-old female tetraplegic 7 years post-injury, with a motor complete C3-C4 spinal lesion. She has no control or sensation of her hands, so attempted hand movements refer to trying to activate the muscles of the hand while imagined hand movements refer to visualizing the movement (without any muscle activity). The study was approved by the California Institute of Technology, Casa Colina Centers for Rehabilitation, and University of

California, Los Angeles Institutional Review Boards. We obtained informed consent after explaining the objectives of the study and the possible risks involved.

4.2.2 Experimental Setup

Experimental sessions were performed at the Casa Colina Centers for Rehabilitation. Tasks were performed in a similar setup as in (Zhang et al., 2017), with NS seated in her motorized wheel chair in a dim room. Tasks were presented on a 27-inch LCD monitor occupying approximately 40 degrees of visual angle, with stimulus presentation controlled using Psychophysics Toolbox (Brainard 1997) and MATLAB. No fixation was required or enforced.

4.2.3 Experimental Design

We used a one-dimensional point-to-point control paradigm as the BMI control task. The participant was instructed to control a cursor to move to the instructed target. To control the cursor, neural signals from imagined or attempted hand squeezes of the left or right hand were decoded into an upward velocity signal, such that squeezing the hand would create a force that "pushed" the cursor upwards, and relaxing the hand would remove the force and allow the cursor to "fall" downwards (see Neural Decoder for more details on the decoder specifications). Thus, NS could control the vertical position of a cursor by keeping her hand squeezed to push the cursor up and keeping her hand relaxed to move the cursor down. The cursor was bounded by the edges of the screen. We used this relatively simple task to allow enough time for data collection for all four conditions within a single session.

The task was split into a training step (collecting data for the decoder to train on), and a testing step (evaluating the decoder's performance during online control). Separate training and testing steps were run for each of the four movement conditions (imagined/attempted left/right hand squeezes). We instructed NS which movement condition to use prior to each training/online control run. For the training step, the cursor moved up and down automatically between two points, while NS was instructed to squeeze (as per the movement condition) in accordance with the direction of the cursor's movement (i.e., squeezing when it was moving up, and relaxing when it was moving down). Targets alternated between the two points, resulting in NS having to alternate between squeezes and releases consecutively. The computer-controlled trajectory of the cursor was determined by a linear quadratic regulator that was calibrated to perform point-to-point movements to the target in a naturalistic manner, reaching the target in ~750ms (Aflalo et al. 2015). Following movement, the cursor would rest on the target for the remaining trial duration (3.3 seconds total). We ran 32 trials per movement condition, with each trial composed of a squeeze and release phase.

For the testing step, the patient was cued to move the cursor between 3 points oriented vertically in a center-out paradigm. In each trial, NS was given 6 seconds to move to the target. For each point-to-point movement, if the allotted 6 seconds elapsed, then a secondary assist was activated, bringing the cursor to the target along the ideal trajectory computed above. This was done so that the initial distance between the cursor's starting position and the target position was constant for each point-to-point movement and independent of the success/failure of the previous point-to-point movement. Between trials, the cursor was held constant at the center for 2 seconds without any target being presented (the intertrial-interval, ITI). Note that NS's motor intentions could only affect the cursor position when a target was presented. They

could not affect the cursor position during the ITI, allowing her to relax during the ITI without it affecting the cursor position. Following the ITI, NS could then continue relaxing to allow the cursor to move downwards, or squeeze her hand to make the cursor move upwards. We ran 20 trials total for a single movement condition at a time. This equated to 40 point-to-point cursor movements (each trial split into a center-to-target and a target-to-center movement)—20 where the cursor was moving upwards and 20 where the cursor was moving downwards. The experiment was run on 8 separate days, with 593 units recorded total (assuming independent populations between days, 74.13±2.9 units per day). We tested all four movement conditions each day, with the order of the movement conditions changed each day to avoid potential order effects that could cause performance differences (e.g., performance being better for early runs than late runs).

In the primary task design, each training run for a condition was immediately followed by online control for the same condition (see Figure 5E). This prevented us from disambiguating whether some results were driven by the conditions being closer in time or by the conditions being more similar/matched. For example, a consistent temporal component in the neural data (e.g., a baseline firing rate drift) could cause consecutive pairs of runs to have more similar neural representations than pairs further apart in time. This would make our result that the representations are preserved from training to online control (Figure 5AB) ambiguous. The analysis for that result compares different runs of the same condition but differing by run type (e.g., ARH training and ARH online control). In the primary task design, however, those pairs of runs are also always consecutive in time and are therefore confounded. In other words, the result could be a consequence of the training and online control runs always being adjacent in time, independent of whether or not the runs were the same condition.

To address this concern, we also collected data using a secondary task as a control experiment for the above confound. This version was identical to the primary task but with the order of the runs changed such that training and online control runs of the same condition were no longer always temporally adjacent (see example blocks in Figure 5F). For example, in each block of four runs, we interleaved two movement conditions A and B in the order: training for Condition A, training for Condition B, online control for Condition B, online control for Condition A. A direct comparison of the relationship between the first and third runs (pairs marked in red in Figure 5F, mismatched conditions but closer together in time) against the relationship between the first and last runs (pairs marked in blue, matched conditions but farther apart in time) would thus allow us to disambiguate the confound. This experiment was run on 4 separate days, with 2 blocks of four runs per day (8 runs total), and 220 units recorded in total (assuming independent populations between days).

4.2.4 Signal Recording Procedures

Two 96-channel Neuroport arrays (Blackrock Microsystems model numbers 4382 and 4383) were implanted in the putative homologues of area AIP and Brodmann's Area 5d. Preoperative fMRI was used to identify array implant location (Aflalo et al., 2015). Only data recorded from the array implanted in AIP (at Talairach coordinate [-36 lateral, 48 posterior, 53 superior]) were analyzed and presented here. A Neuroport neural signal processor (NSP) amplified, digitized, and recorded the neural activity. The Neuroport System (composed of both the NSP and the arrays) has FDA clearance for acute recordings over a duration < 30 days. For this brain-machine

interface clinical trial, we have received FDA IDE clearance (IDE #G120096, G120287) to implant and record from PPC past that 30 day limit.

During recording full-bandwidth signals were sampled at 30 kHz in the Central software suite (Blackrock Microsystems) and high pass filtered (250 Hz cutoff). We used -3.5 times the root-mean-square as the threshold for action potential detection. Each waveform was composed of 10 samples prior to triggering and 38 samples after, with a total duration of 1.6 ms. Units were sorted by hand at the session prior to any data collection using the Central software suite (Blackrock Microsystems), with only high quality, easily isolable, single units being identified and sorted (26.6 ± 2.58 units specially sorted per day). We wanted to analyze the neural data recorded during a real-time online session as opposed to an offline analysis with intensive spike sorting, better emulating the conditions of a real-time BMI, so we purposefully did not perform any additional offline sorting.

4.2.5 Decoding Procedures

We used a decoder fit on the vertical velocity of the cursor as a linear function of the neural population as a whole. Both sorted units and channels without explicitly sorted units (hereafter referred to altogether as “units”) were recorded and analyzed. Note that the decoder used for control was trained on only one strategy-effector combination at a time, differentiating between the squeeze state and release state for that particular condition.

For features, the velocity decoder used the z-scored firing rates of the units. Only units with a minimum raw firing rate of at least 1 Hz were included (on average 96.3 ± 0.2% of all units

across all sessions). Firing rates were sampled with a 50 ms sampling period and then smoothed by a causal exponential filter with a 1.5 second duration (length of 30 samples) and a smoothing factor of 0.75 to filter out high frequency noise.

For the regression, we used MATLAB's lasso function, which implements elastic net regularization. This method tries to minimize the number of terms in the model based on an elastic net mixing value that trades off between lasso and ridge regression (a mixing value of 1 is lasso regression, while a mixing value of 0 is ridge regression). We used an elastic net mixing value of 0.05 for the model, with the exact lambda (regularization coefficient) determined through cross-validation (across 15 values of lambda).

4.2.6 Statistical Analysis

Comparisons between pairs or multiple distributions were performed using non-parametric one-way ANOVA (Kruskal-Wallis), grouping by the different movement conditions being compared (either all four or in the pairwise case two at a time). Samples were the values computed on a per day basis (Figure 3, Figure 5A), per trial basis (Figure 7AB), per run basis (Figure 5CD, Figure 7C), or per unit basis (Figure 2, Figure 6, Figure 7D).

Values shown in figures were computed from data sets after pooling across all days. Error bars indicate bootstrapped confidence intervals on the pooled data. All such confidence intervals were computed with 2000 bootstrap data samples.

All analyses were performed using MATLAB 2017a.

4.2.7 Unit Selection

Units were pooled across days assuming independent populations. Only units with mean firing rates above 0.5 Hz and with a signal-to-noise ratio above 0.5 were included in the analyses to limit low firing rate and noise effects.

4.2.8 Linear Model Analysis for Single-Unit Characterization

For each individual unit, we fit a linear model on the unit's firing rate for when the hand was released and squeezed. This was done for each movement condition separately (one model fit per condition, 8 models total). The hand was considered to be in a squeeze or release state based on the position of the target relative to the cursor, and which action NS would need to perform to successfully complete the trial. For the time window of neural activity, we wanted to minimize potential feedback corrections and isolate the movement intention signals. Based on single unit event related averages from this and past studies (Zhang et al., 2017), we chose a window of 500 ms to 1500 ms after cue onset to isolate the majority of the neural modulatory activity. The significance value of the fit (p-value of the t statistic for the beta coefficients) was used to determine tuning to the corresponding condition (significant if $p < 0.05$, FDR corrected). We also analyzed the R2 of the fit as a measure of strength of tuning. Additionally, we analyzed the beta coefficient of the model and its cross-validated standard error as another measure of strength of tuning and reliability of tuning.

4.2.9 Degree of Specificity

Prior to assessing any possible performance differences between the movement conditions, we needed to verify that the conditions were differentially represented in our neural population. To

do this, we performed a degree of specificity analysis, characterizing the degree to which units were specific to different levels of a variable in training and online control. We first looked at one level of a variable (e.g., the right hand), and computed the degree of specificity to the levels of the other variable (e.g., specificity to imagine and attempt). The degree of specificity was computed by taking the difference of the absolute values of the relevant beta coefficients (those associated with the two levels being compared) normalized by the sum of the absolute values of the beta coefficients. For example, to compute the degree of specificity to imagine and attempt with the right hand, the equation would be:

$$Degree\ of\ Specificity = \frac{|\beta_{ARH}| - |\beta_{IRH}|}{|\beta_{ARH}| + |\beta_{IRH}|}$$

where β is the beta coefficients from the linear model fit above for the associated movement condition. This degree of specificity analysis was done separately for the training data and the online control data. We only included units tuned to at least one of the conditions being compared ($p < 0.05$, FDR corrected). For each distribution, we performed a two-sided sign test to determine whether the medians of the distributions were significantly different from 0 (i.e., if they were biased to one variable over another).

4.2.10 Maintenance of Single-unit Tuning from Training to Online Control

Performance differences could be caused by differences in how well each movement condition was maintained from training to online control. As a preliminary measure, we first wanted to assess any such differences at the single unit level. We identified the units tuned to each condition during training and online control (as defined from the linear model analysis above). For each condition, we then computed the fraction of units that kept their tuning to the condition. For example, for the attempt right hand (ARH) condition, the fraction was computed

as the number of units tuned to ARH during both training and online control divided by the number of units tuned to ARH during training.

4.2.11 Correlation between Neural Representations

Besides the single-unit measures of how well representations were maintained, we also wanted a population measure of the similarity between representations during training and online control. Using correlation as a measure of similarity, we directly looked at the similarity of the neural representations for each condition separately. We used correlation over other distance measures (such as Euclidean or Mahalanobis distance) because correlation gives a normalized value of the similarity between the representations and is invariant to gross baseline changes across the entire population.

The neural representations were the vector of normalized beta coefficients (one element for each unit). The normalized beta coefficients are the beta coefficients (from the linear models above) normalized by their 95% confidence intervals and thus are a trial average measure of each unit's activity weighted by its trial-to-trial variability.

4.2.12 Comparison of Representations Between Training and Online Control

The above analyses treat each condition separately, looking at how well representations were maintained from training to online control. We also wanted to test how well the structure of the representations of the movement conditions as a whole were preserved going from training to online control (i.e., the relationships between the representations). To do this we performed a cross-decoder analysis, training a linear classifier (linear discriminant analysis, equal diagonal

covariance matrices for each condition) on just the training data to classify the four movement conditions, and then tested the ability of the classifier to generalize to the online control data. The classifier used the modulation of the units' firing rates from the release state to the squeeze state as features. The classifier's cross-validated performance on training data was also computed for comparison. Also, a second classifier was trained on the online control data and tested on the training data (testing generalization from online control representations to training representations). We performed this cross decoding analysis independently for each of the experimental sessions, resulting in a distribution of scores for the cross-validation/generalization scores of each classifier. For each score distribution, we used a one-sided Wilcoxon signed rank test to determine if the scores were significantly above a chance prediction score of 1 out of 4 (0.25, $p < 0.05$).

In order to look at how well the classifiers performed with each condition during generalization, we also computed the confusion matrices for each of the generalization scores. For each trial, we recorded the true condition type as well as the condition type predicted by the classifier. These true- and predicted- condition pairs were counted up and tabulated into a matrix form and then normalized by the number of trials per condition, resulting in the confusion matrix values shown in Figure 5B.

4.2.13 Analysis to Control for Order Effect Confound

In order to determine whether the results of the above analysis were driven by the conditions being matched or the runs of matching conditions also being temporally adjacent, we performed two additional control analyses.

The first analysis was done on the primary task data set. We compared the similarity between neural responses from training to online control when the conditions were matched (and therefore adjacent due to the primary task's design) against the similarity when the conditions were mismatched but still adjacent in time. For example, consider the set of runs in the sequence depicted in Figure 5E. We compared the similarity in the responses of consecutive pairs with matched conditions (pairs marked in blue) against the consecutive pairs with mismatched conditions (pairs marked in red). We measured similarity as the correlation between the beta coefficients of the conditions. The distributions of the correlation values was then compared between the two groups (matched vs mismatched) using a nonparametric one-way ANOVA (Kruskal-Wallis).

The second analysis was done on the secondary task data set. For each block of four runs (see Experimental Paradigm, Figure 5F) we directly compared the correlations between the first and third runs (pairs marked in red, mismatched by condition but closer together in time) against the correlations between the first and last runs (pairs marked in blue, matched by condition but farther apart in time). The distributions of the correlation values were again compared with a nonparametric one-way ANOVA (Kruskal-Wallis).

4.2.14 Neural Performance

In the above analysis, we used tuning as a simple single-unit measure of how well a representation was maintained. However, one drawback of this method is that it makes a binary determination of whether or not a unit is tuned. It is possible for a unit to be classified as tuned

during both training and online control but with different levels of significance. For example, a unit could be significantly modulated by a condition during both training and online control but with significantly different amplitudes.

To account for this possibility, we used the performance of individual units during online control as a continuous measure of how well units maintained their tuning. For each unit, we looked at the direction of the unit's "force" on the cursor within a time window (e.g., the above 500 ms to 1500 ms used in the above linear analysis). For the given time window, the neural activity was projected through the corresponding decoder weight (from a decoder trained on the training data) into a "force" acting on the cursor. The sign (direction) of the average force across the time window indicated whether the unit was pushing the cursor up or down on average during that interval. We then compared the direction of this force to the direction the cursor would need to move in to reach the target. In other words, we looked at whether the unit was pushing the cursor in the correct direction on average. This was computed on a trial-by-trial basis for the selected time window. The neural performance of an individual unit was then defined as the fraction of trials where the unit was pushing the cursor in the correct direction.

When aggregating these single unit measures, we used a weighted average of neural performance values, with weights taken from the corresponding population decoder used for online control. This allows us to directly measure the generalization performance of each unit in the context of its effect on online control performance. In other words, the weighted average is reflective of the performance of a decoder using just that subset of units. While comparing firing rates between training and online control would also be informative of how well units

maintained their tuning, it would not account for the units' actual effects on online control. Note that for the decoder weights we used the weights from the population decoder as opposed to the single-unit beta coefficients.

4.2.15 Behavioral Performance Metrics

We used two metrics to assess behavioral control performance. The first metric was simply the number of successful trials divided by the total number of trials (success rate). A trial was considered successful if the cursor reached the designated target and was held there for 1 second within the allotted 6 seconds (under NS's own control). The second metric was the time required to reach success (not including the 1 second hold time), looking only at successful trials.

To get a sense of what constituted a "good" success rate, we used real neural data to simulate the chance of the cursor reaching the "wrong"target. For each online control trial, we defined the "wrong" target of the simulated trial as the target in the opposite direction from the true target. Using the neural data and decoder from that trial, we simulated the trajectory of the target. We then marked the simulated trial as a "success" if the simulated trajectory reached the simulated target and was held there for 1 second (the same criteria as was used on the real online control trials). The cursor bounds keeping the cursor on screen were also translated so that the simulated trials had bounds symmetrical to the real trial. For example, consider a true trial where the cursor moved from the top target down to the center and with a boundary directly above the starting position, preventing it from moving upward too much. The corresponding simulated trial would require the cursor to move from the top target to a mock

target above the top target and with a boundary directly below the starting position, preventing it from moving downward too much. The mock target would be placed such that the starting top target would be equidistant from the mock target and the true center target, requiring it to travel the same distance in the simulated trial as in the true trial. This was done so that the effects of the boundaries on performance would be replicated in the simulation as well. This simulation method allows us to determine whether our observed control performance was significant, i.e., whether NS had control.

4.3 Results

4.3.1 Representations during Training and Online Control

In the first part of this study, we investigate how similarly individual movement conditions involving various effectors and cognitive strategies are represented between training and online control, as well as how the *structure* of these representations (i.e., relationship between the representations) is maintained.

Recording from AIP of a female, C3/C4 tetraplegic participant (NS), we compared neural responses of four movement conditions during the calibration ("training") and online control steps of the 1D BMI control task (Figure 1AB). The four movement conditions tested were: attempted right hand movements (ARH), imagined right hand movements (IRH), attempted left hand movements (ALH), and imagined left hand movements (ILH). NS controlled the cursor by squeezing her hand to "push" the cursor up, and relaxing her hand to let the cursor "fall". During each block of training and control we instructed NS to use a particular movement condition (e.g., to attempt right hand squeezes). We trained the decoder using a point-to-point

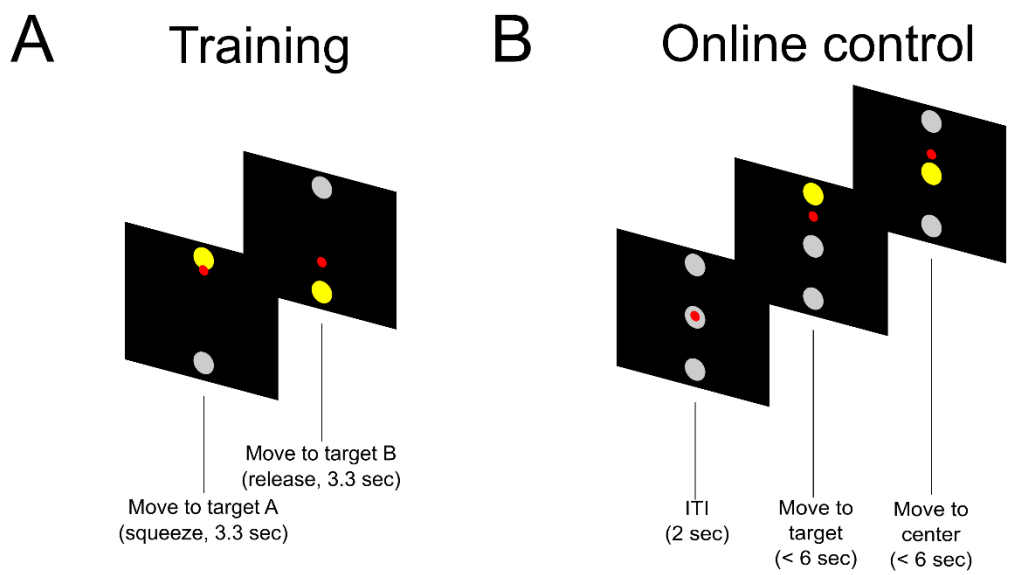

Figure 4.1. Experimental Paradigm

(A) Training task. The small red circle is the cursor, the gray circles are the possible targets, and the yellow circle is the target for the specific trial. (B) Online control task.

paradigm (Figure 1A), with the cursor alternating between two points under computer assistance. During online control, the task was changed to a one-dimensional center-out task (Figure 1B) while the decoder (trained on the corresponding strategy and effector) decoded NS's intentions. Each trial started with the cursor fixed at the center position for 2 seconds while the cued target was highlighted. Without any computer assistance, NS had 6 seconds to control the cursor to the cued target and hold it there before computer assistance was enabled to help complete the trial. See Methods for more details about the decoder and experimental design.

We first wanted to verify that the four movement conditions were represented in the recorded population, during both training and online control, by looking at the percent of the population tuned to each condition. For each unit, movement condition, and run type separately, we fit a linear model to the unit's firing rate modulation between the "release" to "squeeze" hand states (see Methods for more details). A unit was considered tuned if the slope of the beta coefficient

of the model was significantly different from zero ($p < 0.05$, FDR corrected). A significant fraction of the population was tuned to each of the conditions (Figure 2A), indicating that they were indeed represented.

We used a specificity analysis to examine the degree of overlap between the populations representing each condition (similar to the one performed in (Zhang et al., 2017)). If populations were non-overlapping, we would expect to see most units having high specificity values. On the other hand, if populations were overlapping, the distributions would be more uniformly distributed. Figures 2B-E show the distribution of the specificity values focusing within one level of a variable at a time. For example, Figure 2C looks at the degree of specificity of the units to attempt and imagine only for conditions involving the right hand. The degree of specificity was computed as the difference in the absolute values of the beta coefficients (from the linear model), normalized by the sum of the absolute values of the beta coefficients. In Figure 2C, a value of 1 would correspond to a unit activated only by attempt and not at all for imagine, a value of -1 would correspond to a unit activated only by imagine and not attempt, and a value of 0 would correspond to a unit activated similarly by both. Units not significantly tuned to either were excluded from the analysis. We computed the distributions separately for the training data (in blue) and the online control data (in orange). All computed distributions indicated partially overlapping populations, with some units highly specific to a condition and other units equally responsive to both.

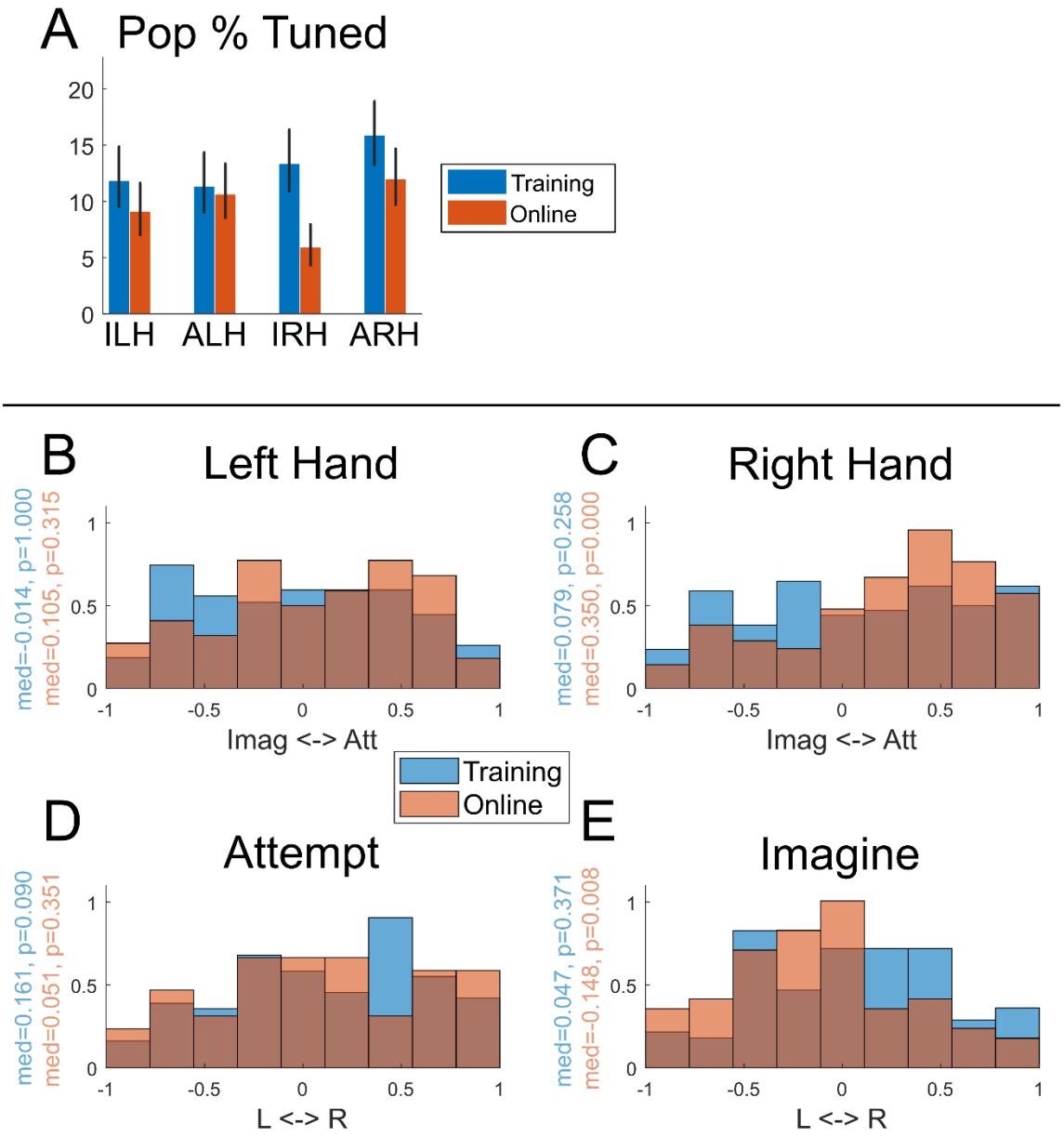

Figure 4.2. Population Tuning to the Movement Conditions

(A) Percent of units tuned to each movement condition (bootstrap 95% CI, $p < 0.05$, FDR corrected). See Table 1 for an ANOVA of the values. (B) Degree of specificity showing distribution of how much units exclusively code ILH or ALH. Distribution during training shown in blue and distribution during online control shown in orange. For each distribution, the median and the probability the median is different from 0 (two-sided sign test) are shown in their corresponding colors. (C) Similar to (B) but for IRH and ARH. (D) Similar to (B) but for IRH and ARH. (E) Similar to (B) but for IRH and ARH.

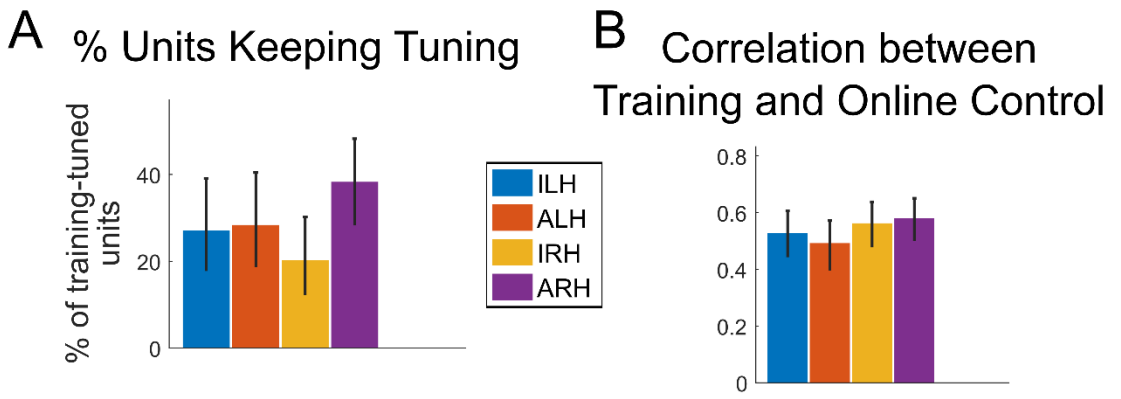

Figure 4.3. Consistency of Representations between Training and Online Control

(A) Percent of units (out of those tuned during training) that were tuned during both training and online control (bootstrap 95% CI). See Table 3 for an ANOVA of the values. (B) Correlation between movement representations during training and online control (bootstrap 95% CI).

Given that all four tested movement conditions were represented in different but partially overlapping populations, we first wanted to see whether there were differences in how well the subpopulations of units tuned to each of the movement conditions kept their tuning between training and online control. We looked at the units tuned to each condition during training and asked what fraction of these tuned units were also tuned to the condition during online control. The results are shown in Figure 3A. There was no significant difference between conditions in the fraction of units that kept their tuning ($\chi^2(3, N = 32) = 5.80$, $p = 0.12$, Kruskal-Wallis).

We also examined how well the representations were maintained at a population level. Using correlation as a measure of similarity, we correlated the neural representations of each condition during training to its corresponding representation during online control (see Methods for more details). There were no significant differences in the similarity between neural representations going from training to online control (Figure 3B, $\chi^2(3, N = 32) = 6.59$, $p = 0.087$, Kruskal-Wallis).

The above analyses in Figure 3 suggest that the representations of all four conditions are largely preserved between training and online control. The comparable level of maintenance between each of the conditions also suggests that the *structure* of the representations itself is maintained. However, this is not an obvious result. The above analyses treat each condition separately and do not look at the how the relationship between the conditions changes between training and online control. In Figure 4A-C we show three main possible configurations of the structure of the representations going from training to online control. (1) The structure of the representations is maintained and consistent between training and online control, as the above results seem to suggest (structure maintained, Figure 4A); (2) all four conditions are differently represented during online control but in a different structure than during training (structure different, Figure 4B); and (3) the representations of the four conditions collapse into a single representation that is invariant to which of the four conditions is being used, such as in a pure intention or goal signal (structure collapsed, Figure 4C).

To more directly adjudicate between the three configurations, we performed a cross decoding analysis to test how well the representations of the four conditions generalize across the run types (i.e., across training and online control). We trained a linear classifier on the training data to classify between the four movement conditions and tested it on the training data (cross-validated performance) and the online control data. Conversely, we also trained a classifier on the online control data and tested it on the online control data (cross-validated performance) and the training data (see Methods for more details).

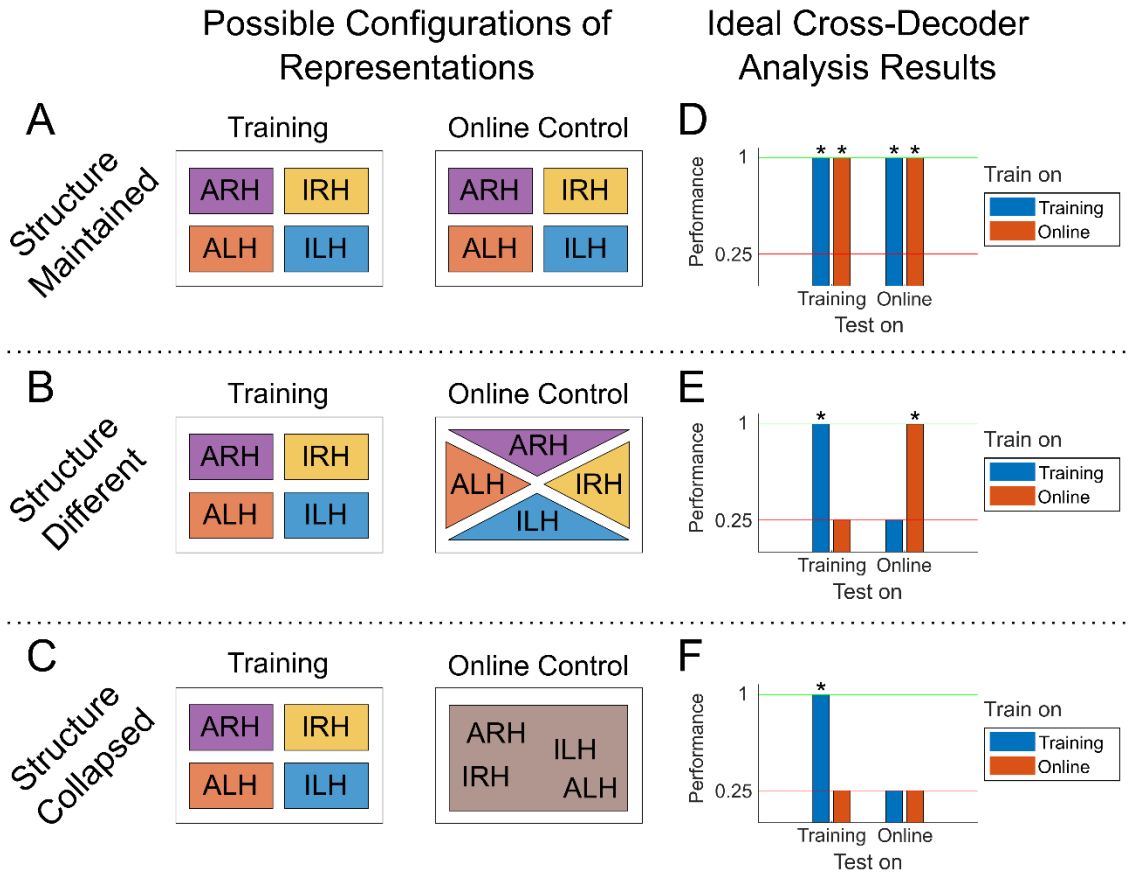

Figure 4.4. Possible Configurations and Corresponding Expected Analysis Results

(A-C) Schematics for different possibilities in how the structure of the representations compares between training and online control. (A) Schematic for the "structure maintained" case where the structure is consistent between training (left) and online control (right). Representations of the four movement conditions are separable during both training and online control, and in the same structure (i.e., the same configuration, as represented by the consistent placement of the conditions). (B) Schematic for the "structure different" case where the movement conditions are separable during both training (left) and online control (right) but with different structures (i.e., different configurations). (C) Schematic for the "structure collapsed" case where the movement conditions are separable during training only (left) and collapse into a single representation (as represented by the conditions being no longer separable in the online control case, right). (D-F) Ideal expected result from cross-decoding analyses if the data follows the different schematics in Figure 4A-C. See Results for detailed explanation of colors and bars. Red lines represent chance performance (0.25). Performances significantly above chance are marked. (D) Ideal expected result in the "structure maintained" case of Figure 4A. (E) Ideal expected result in the "structure different" case. (F) Ideal expected result in the "structure collapsed" case.

The results of such an analysis can directly clarify which of the above three configurations fits our data better. Significant cross-validated classifier performance within the trained-on run type would mean that the four movement conditions are differently represented during that run

type, while chance-level performance would mean that they are not easily distinguishable from each other. Significant performance of the classifier on the not-trained-on run type (i.e., generalization performance) would mean that the representations are preserved between the run types, while chance-level performance would mean they are different.

Idealized example results corresponding to each of the three configurations are shown in Figures 4D-F. If the structure is maintained (Figure 4A), we would expect results similar to those in Figure 4D. The training classifier in this case not only performs well within the trained-on training data (right blue bar, cross-validated performance), but also performs well with the not-trained-on online control data (left blue bar). Similarly, the online control classifier performs well both with the trained-on online control data (right red bar, cross-validated performance) and with the not-trained-on training data set (left red bar). This bidirectional generalization indicates that the four conditions are differently represented in both training and online control and that these representations are similar between the two run types. On the other hand, if the structure is different (Figure 4B), then we would expect results similar to those in Figure 4E, where both cross-validated performances are significant, but generalization performance is only chance-level. Finally, if the structure collapsed during online control (Figure 4C), we would expect cross-validated performance for the training classifier to be significant, but cross-validated performance for the online control classifier to be chance-level (representations not different during online control) along with the generalization performance (Figure 4F).

With our data, we found that the classifiers generalized well from training to online control (and vice versa), performing significantly above chance (Figure 5A). This suggests that the structure of

the representations of the four movement conditions is maintained and meaningful across the two run types (structure maintained, Figure 4A). Looking further at the confusion matrices shows how well the classifiers generalized for each condition. We found that the representations of all four conditions tended to generalize equally well (Figure 5B).

In our experimental procedure, online control runs always occurred directly after the training run of the corresponding condition. As a result, the above cross-decoding analysis has a potential order effect confound. The cross-decoding generalization could simply be due to run types of the same condition type (e.g., ARH training and ARH online control) always being temporally adjacent to each other. For example, if the neural data had a temporal component that changed consistently independent of the movement conditions (e.g., a baseline mean firing rate drift causing representations to change), then we would also expect to find the maintained structure found above. The maintained structure would be a consequence of the neural representations being recorded closer together in time (temporally adjacent), independent of whether or not they were matched by condition.

To control for this possibility, we compared the correlations of the neural representations between training and online control (pairs marked in blue in Figure 5E, temporally adjacent, "matched" by condition), against the correlations between the other temporally adjacent condition combinations (pairs marked in red, temporally adjacent, but "mismatched" by condition). In other words, we compared the similarity of the neural representations between each training run and its *following* online control run (matched and adjacent) against the similarity of each training run and its *preceding* online control run (mismatched and adjacent).

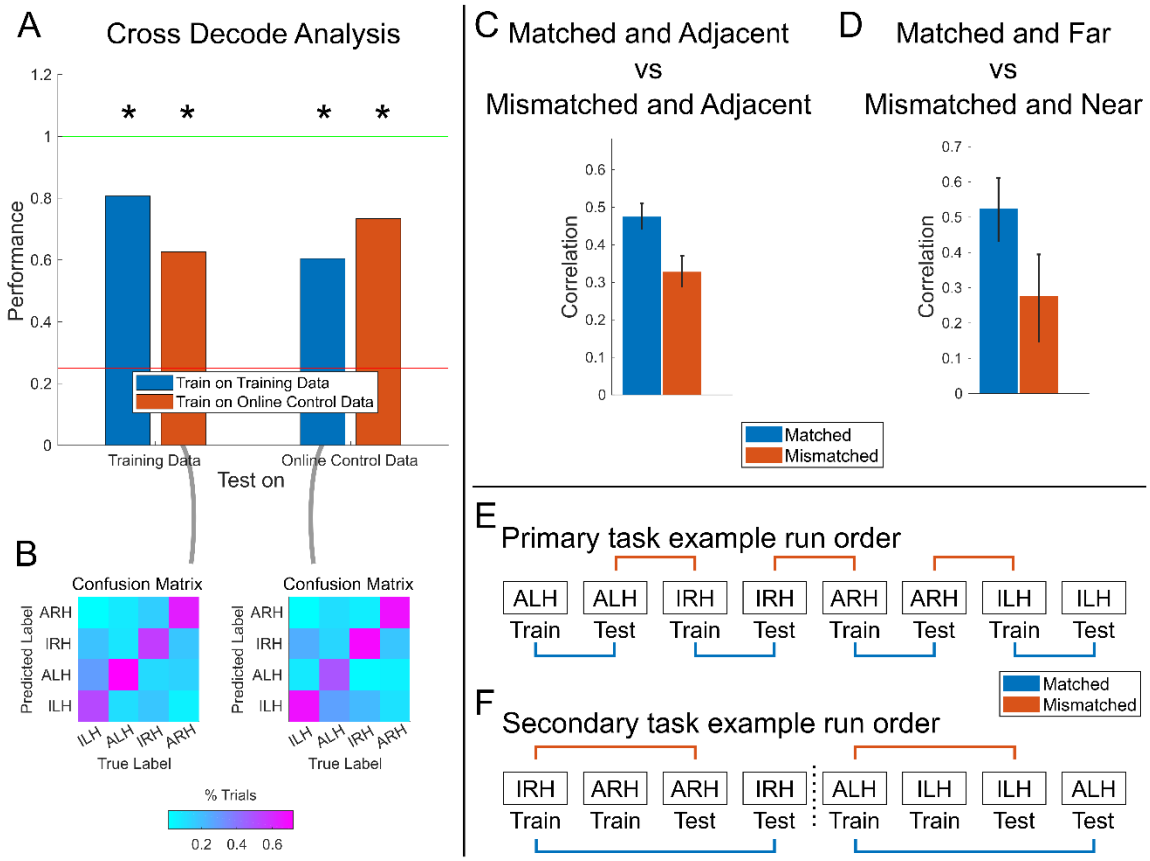

Figure 4.5. Maintenance of the Structure of the Representations

(A) Results of the cross decode analysis performed on our data, presented as in Figure 4D-F. Performances significantly above chance are marked (one-sided Wilcoxon signed rank test, $p < 0.05$, see Methods for more details). (B) Confusion matrices showing classifier predictions when generalizing from one run type to the other, shown as the percent of trials per condition. Columns are the true condition labels and rows are the predicted labels. Left matrix corresponds to the classifier trained on the online control data and tested on the training data (Figure 5A, left red bar). Right matrix corresponds to the classifier trained on the training data and tested on the online control data (Figure 5A, right blue bar). (C) Correlation between neural representations of pairs of runs where the runs were adjacent in time and matched in condition (blue), compared to the correlation between pairs adjacent in time mismatched in condition (red). Error bars are 95% bootstrapped confidence intervals. See Methods for more details. (D) Correlation between neural representations of pairs of runs matched by condition but farther apart in time (blue) compared to pairs mismatched by condition but closer together in time (red). Error bars are 95% bootstrapped confidence intervals. See Methods for more details. (E) Example set of runs from a single session for the primary task paradigm (see Methods). Pairs marked in blue are matched by condition and adjacent in time while pairs marked in red are mismatched in condition but still adjacent in time. (F) Example of two blocks of runs (4 runs per block) for the secondary task paradigm used to control for an order effect (see Methods). Pairs marked in blue are matched by condition but farther apart in time while pairs in red are mismatched by condition but closer in time.

Despite all examined pairs of data sets being temporally adjacent, the correlations between pairs matched by condition were significantly higher than those not matched by condition (Figure 5C, $\chi^2(1, N = 56) = 20.88$, $p = 4.89e\text{-}6$, Kruskal-Wallis).

We also controlled for this confound by collecting several secondary data sets where not all training and online control runs of matching conditions were temporally adjacent. Instead of each training run being immediately followed by an online control run of the same condition, we used the following order in each "block" of runs: training for Condition A, training for Condition B, online control for Condition B, online control for Condition A—where the Conditions A and B were each selected from the four movement conditions (see Figure 5F for example blocks of runs). This ordering allowed us to directly test whether the correlations were driven more by the matching of conditions or by the temporal closeness of the conditions. We compared the correlations between the neural representations (beta values) of the first and last runs of each block (pairs marked in blue in Figure 5F, same conditions, but farther apart in time) against the correlations between the neural representations of the first and third runs (pairs marked in red, different conditions, but closer together in time). The neural representations were more similar when the conditions were matching, rather than when the conditions were closer together in time (Figure 5D, $\chi^2(1, N = 16) = 9.28$, $p = 0.0023$, Kruskal-Wallis). Altogether, the results of Figure 5CD suggest that the "structure maintained" configuration found in Figure 5A is not a result of the fact that the conditions were temporally adjacent. Rather, the result was driven by the actual matching of the conditions.

The above results show that the structure of the representations of the different movements is relatively consistent between training and online control, with significant generalization in the organization from one run type to the other. However, the generalization is not perfect, with the generalization performance still lower than the cross-validated performance. Figure 5B already shows that the representations generalize equally well for each of the movement

conditions, albeit imperfectly, meaning there is no single movement condition that causes the generalization performance to drop.

The drop in generalization performance could also be due to a specific subset of units generalizing poorly, rather than all units (regardless of tuning preference) generalizing imperfectly. Thus, we next asked if there was a systematic difference in how well specific units generalized based on their tuning preference. For example, do imagine-specific and attempt-specific units both maintain their specificity equally well or does one type generalize to online control better?

To answer this question, we focused within one level of a variable at a time (e.g., the right hand), categorized units by their specificity to the levels of the other variable during training (e.g., only tuned to attempt, only tuned to imagine, or tuned to both), and then assessed their performance during online control (Figure 6). For example, Figure 6B looks only at units for conditions involving the right hand. Units were identified as attempt-specific, imagine-specific, or non-specific based on their tuning during training (i.e., tuned to attempt only, tuned to imagine only, or tuned to both). We then evaluated the average performance of these groups of units during online control. Individual unit performance was determined by whether the contribution of that unit's activity to the decoder positively or negatively affected the cursor velocity (i.e., whether the activity was pushing the cursor towards the target correctly). This

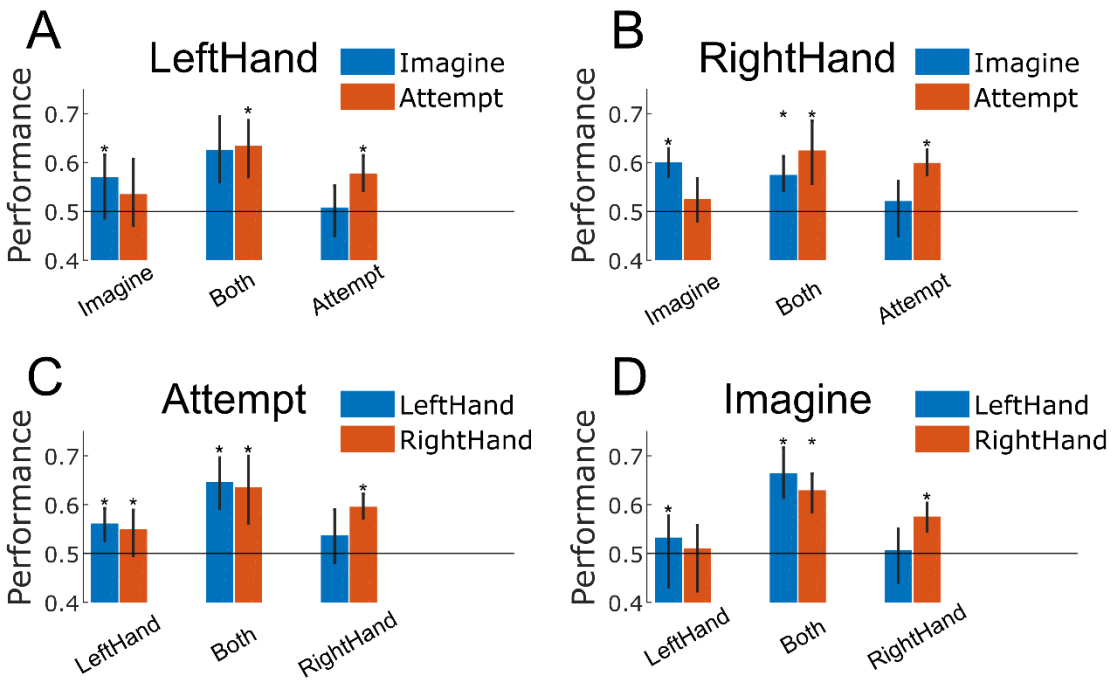

Figure 4.6. Maintenance of Representations split by Tuning Preference

(A) Average single unit performance (weighted by the corresponding decoder weights) for imagined/attempted left handed movements (bootstrap 95% CI). Units are grouped by tuning only to attempted movements, tuning only to imagined movements, and tuning to both. Performance was evaluated for imagined left handed movements (blue bars) and attempted left hand movements (red bars). Performances significantly above chance (one-sided sign test, $p < 0.05$, FDR corrected) are marked and chance performance is marked by the solid line. (B) Similar to (A) but for right handed movements. (C) Average single unit performance (weighted by the corresponding decoder weights) for left/right handed movements using the attempt strategy. Units are grouped by specificity of tuning to the left or right hand, with performance evaluated during left- and right-handed movements (blue and red bars, respectively). Significant performances are marked. (D) Similar to (C) but for movements using the imagine strategy.

measure of performance is more continuous than the tuning measure used in Figure 3A that only looked at the percentage of units that kept their tuning, allowing us to take into account the behavior of units that keep their tuning but not as strongly (see Neural Performance in Methods for more details). Furthermore, the aggregate neural performance of the individual units in Figure 6 is a weighted average, with the weights from the corresponding decoder used for online control. This allows us to account for each unit's effect on online control performance, granting units with a greater contribution more influence. Thus, the weighted average is reflective of the performance of a decoder using only that subset of units. In general, units that were specific to one condition maintained their specificity in terms of performance during

online control. This was true for each of the four conditions. Likewise, units that were non-specific between the compared conditions ("Both") performed equally well with either condition, maintaining that non-specificity. For example, units that were tuned to ARH and not IRH during training performed above chance with ARH and not with IRH and vice-versa (Figure 6B). Similarly, units that were responsive to both ARH and IRH performed comparably well with both ARH and IRH during online control (Figure 6B). These results indicate that the tuning preference of a specific unit does not affect how well it will generalize from training to online control and that there is no specific functional variable that generalizes better than another. These results also further emphasize that the single unit tuning preferences observed during training are largely maintained and meaningful during online control, consistent with our population results above.

4.3.2 Comparison of movement conditions and online control performance

In the second part of this study, we focus on whether all the tested movement conditions are feasible for online BMI control, the degree to which performance differs between the movement conditions, and the possible causes of the performance differences.

We compared the performance of the movement conditions individually, looking at both trial success rate (fraction of trials where NS successfully moved the cursor to the target within the allotted 6 seconds) and the time to successful trial completion. We found that all four combinations of strategy and effector resulted in significant control performance (Figure 7A). Interestingly, the ARH condition performed significantly better than the other three (IRH, ALH, ILH). This was true when using both trial success rate as a measure of performance (Figure 7A,

$\chi^2(3, N = 1280) = 19.06$, $p = 2.66e\text{-}4$, Kruskal-Wallis) and time to trial completion of the successful trials (Figure 7B, $\chi^2(3, N = 994) = 16.43$, $p = 9.24e\text{-}4$, Kruskal-Wallis).

Hypothetically, there are several possible explanations for why ARH performed better than the other conditions did. First, the representation of ARH during training might be more similar to its corresponding representation during online control, either at a single unit level or a population level. This would lead to the decoder trained on the ARH training data performing better during online control. Second, more units could be tuned to ARH, resulting in a larger signal and thus better control performance. Finally, the ARH tuned units might be more strongly tuned than the other conditions, having more information content on a unit-by-unit basis, also leading to a larger signal.

In regards to the first possibility, our above results already show that there are no differences in how well the different movement representations are maintained. Not only are each of the representations equally well maintained, but their structure is maintained, too (Figures 3, 5).

Thus, we first sought to explain the performance differences simply as a function of the number of units used for each decoder. Although the number of units recorded by each of the four conditions' decoders was not significantly different, there were differences in how many of those units were actually tuned. Focusing on the training data, a test of equal tuning percentages for all four units only trended towards significance (Figure 2A, $\chi^2(3, N = 2372) =$

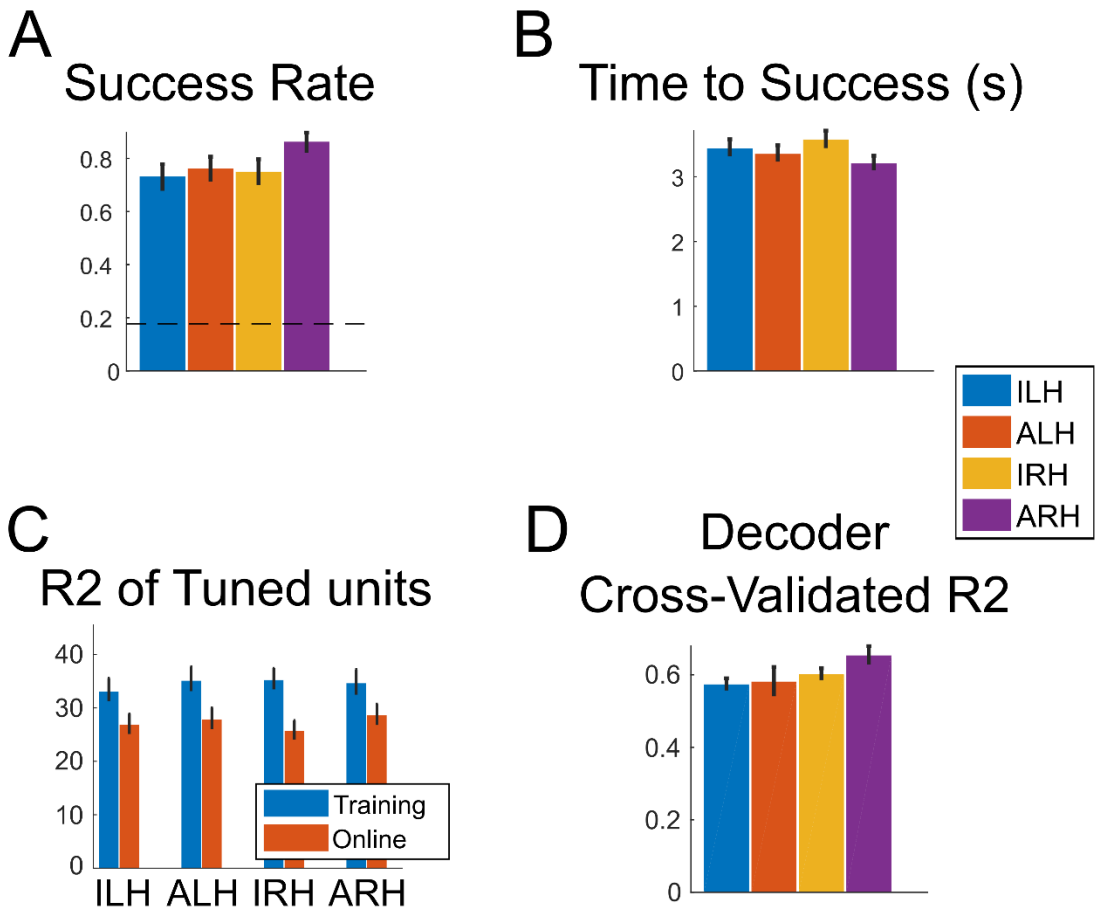

Figure 4.7. Online Control Performance

(A) Performance of each movement condition, measured as the fraction of successful trials (bootstrap 95% CI). Dashed line indicates simulated chance performance (see Methods). (B) Performance of each movement condition, measured as the mean duration of successful trials (bootstrap 95% CI). (C) Mean R2 of units tuned to each movement condition from Figure 2A (bootstrap 95% CI). See Table 2 for an ANOVA of the values. (D) Cross-validated R2 of the decoder used for online control, trained on the training data for each condition (bootstrap 95% CI). Cross-validated R2 was computed for each condition and session separately.

5.86, p = 0.12, Kruskal-Wallis on the significance values of the linear model fits). However, a test of ARH compared to all other conditions showed a significant difference ($\chi^2(1, N = 2372) = 4.35$, p = 0.037, Kruskal-Wallis), consistent with the observed performance differences.

We next asked whether there were any differences in the strength of tuning of the units tuned to each condition. The performance differences could be driven not only by the greater number

of ARH tuned units, but also the ARH tuned units being more strongly tuned, having greater information content and being more reliable. We used the R2 of the linear model fits computed previously as a measure of the strength of tuning. Once again focusing on only the training data, while ARH tended to have more tuned units, on average, ARH tuned units were not any more strongly tuned than units tuned to other conditions (Figure 7C, $\chi^2(3, N = 333) = 1.99$, $p = 0.58$, Kruskal-Wallis). These results suggest that on a unit-by-unit basis, there is nothing qualitatively special about the units tuned to ARH compared to the units tuned to other conditions.

Previously, we found that the correlations of the movement conditions between training and online control were also all comparable (Figure 3B) and that the structure of the representations is largely maintained between training and online control (Figure 5). In light of these results, it makes sense that any performance differences between the conditions existing during training might carry over to online control. In other words, it should be possible to predict the online control performance trends based on solely looking at the training data. To test this, we examined the cross-validated R2 of each movement condition's decoder (decoders trained on data from the training runs, Figure 7D). ARH had a significantly higher cross-validated R2 than the other conditions ($\chi^2(3, N = 32) = 10.69$, $p = 0.014$, Kruskal-Wallis). This is the same trend as found in our performance measures (Figure 7AB). Because of the "structure maintained" result, the properties of the movement conditions during training were indeed preserved in online control.

4.4 Discussion

In this study, we recorded from human AIP of a tetraplegic participant and investigated the degree to which mixed representations of attempted or imagined movements of the left or right hand decoded maintained their representations from training to closed-loop cortical control. Our results show that the structure of the representations were all maintained between training and online control. Furthermore, the different tested effectors and cognitive strategies could all be used for control, with performance differences primarily due to differences in the number of units tuned to each movement condition during training.

4.4.1 Consistency of representations from training to online control

A recent study out of our lab has found partially mixed representations in human AIP, with strategy and body part variables mixed and functionally segregated by body part (Zhang et al., 2017). This recent study focused only on the coding of variables in the absence of any closed-loop BMI control (i.e., representations during "training" only). The current study, however, examines these representations not just during training but also during online control, investigating the degree to which the representations change from one to the other.

Unlike our recent study, in this study we focused only on movements below the level of injury, looking at how hand movements are represented and not shoulder movements. This study was conducted as part of a BMI clinical trial with a focus on using body parts below the level of injury for online control. Furthermore, this was done in consideration of experimental time constraints, as a full set of 8 movement conditions (attempt/imagine left/right hand/shoulder movements) with both training and online control would have been too taxing

on NS, resulting in overall decreased task performance. Considering that we are recording from the same brain area as in (Zhang et al., 2017), we would expect results to be largely consistent between the two studies had we also included shoulder movement conditions (e.g., finding functional segregation by body part).

Consistent with (Zhang et al., 2017), we found that a significant fraction of our recorded neural population encoded the tested movement conditions during training (Figure 2A). Furthermore, populations tuned to each of the variables were partially overlapping, with units having varying degrees of specificity to one variable over another (e.g., attempt over imagine, Figures 2B-E), also consistent with the past study.

In addition to reproducing our previous study's results, we found that these representations are largely maintained between training and online control. Not only did individual units tend to keep their tuning and specificity between training and online control (Figure 3A, Figure 6), but the population representations as a whole also stayed relatively similar between the two run types (Figure 3B). Remarkably, despite the strategy and body side variables having significantly correlated neural representations and thus being more difficult to decode (Zhang et al., 2017), the representations of each were relatively well preserved between training and online control and were still distinguishable by our decoders.

The relative maintenance of representations from training to online control is consistent with the ability to use recordings from AIP for brain control (Aflalo et al., 2015; Revechkis et al.,

2014). While there can certainly be tuning changes in some of the units when moving to online control (Chase et al., 2009; Cunningham et al., 2011), a percent of the population must be relatively consistent in its behavior for a decoder trained on the training data to generalize and perform online.

Furthermore, we find that not only are the representations of the tested movement conditions largely maintained, but the *structure* of their representations (i.e., the relationship between the representations) is maintained (Figure 5AB). Past studies involving BMI control have focused on one effector at a time or have not compared the representations of different effectors between training and online control, and thus have failed to shed light on this question. It is not obvious a priori that the representations of all the different movement conditions would be preserved between the run types. In particular, an alternative possibility is that the representations are no longer distinct during online control and instead collapse into an effector-independent representation used for BMI control. Some studies have found goal or intention signals in PPC but these signals are generally tied to the effector. For instance, that the parietal reach region is activated by reaching to goals but not saccading to them, and that the reverse is true for the lateral intraparietal area (LIP) (Batista et al., 1999; Cui and Andersen, 2007; Snyder et al., 1997). The results of Figure 5AB further support the idea of effector-dependent intention signals. The type of movement condition is still meaningful during online control, with their representations not only still distinct during online control, but also in a way that the relationships between the representations are consistent between training and online control. On the other hand, it has been proposed that LIP is related to saliency of stimuli independent of effectors (Bisley and Goldberg, 2010). The finding that there is strong effector specificity, under two very different

behavioral conditions of training and online control, indicates that in humans as in monkeys, the intended movement of specific effectors is a guiding feature of the population structure in PPC.

The maintenance of the distinctions between the movement conditions and structure of their representations also suggests that it is possible to control multiple effectors recording from a single brain area. In this study, we were able to successfully decode not just the onset of the movements (i.e., squeeze and release), but also the body side and cognitive strategy employed in the movement. To the best of our knowledge, past studies have only looked at BMI control using multiple effectors (e.g., bimanual control) in the context of multiple brain areas (Ifft et al., 2013).

Some studies propose that the only difference between imagining and attempting/executing a movement is the overt motor movement itself (Jeannerod, 1995). This would suggest that the representations of an imagined or attempted movement would be indistinguishable at a neural level in the absence of overt movement. Similar to our results from (Zhang et al., 2017), we find that imagined and attempted movements are in fact distinguishable from each other at a neural level even when there are no overt movements. Furthermore, in this current study, we find that these differences persist and are consistent even when the movements are “overt” and directly control an effector (i.e., in BMI online control). Thus, we find that our results are inconsistent with this theory and that imagined movements are not just a subset of attempted movements at a neural level.

4.4.2 Performance differences of different effectors during online control

In this study we assessed the online control performance of four movement conditions (attempted/imagined movements of the left/right hand) and found that all performed significantly above chance (Figure 7AB). Despite the representations of each of the conditions being similarly well maintained between training and online control (Figure 3), attempted movements of the right hand (ARH) performed significantly better than the other movement conditions. The information content of individual units did not significantly differ based on the movement condition, either (Figure 7C).

The primary difference in the representations of ARH compared to the others was the greater proportion of tuned units (Figure 2A) found in the training data. In this context, the maintenance of the structure of the representations makes these differences carry over to online control. Since the relationship between the representations does not change, the performance differences observed during training would not significantly change during online control either. In other words, the maintenance of the structure makes it possible to predict relative online control performance based on offline training data, as demonstrated by the similarity in the trends between the decoder cross-validated R2 (Figure 7D) and the online control performance (Figure 7AB).

The greater proportion of units tuned to ARH is also consistent with our array recording location. The array is located in left AIP, a region traditionally thought to encode grasp information of the contralateral limb more specifically (Chang et al., 2008; Murata et al., 2000). A preference for the right hand would be plausible in this context.

Furthermore, some studies on BMI control using EEG and other recording technologies found attempted movements to perform better than imagined movements (Blokland et al., 2014; López-Larraz et al., 2012). This is consistent with our finding that the attempt strategy performed better than the imagine strategy in the right hand (Figure 7AB) as well as the bias towards attempted over imagined right hand movements in the degree of specificity of the individual units during online control (Figure 2C).

This study was part of a clinical trial composed of a variety of experimental tasks involving BMI control beyond those presented here. Most of those studies involved attempted movements of the right hand. Thus, NS has had significantly more practice using attempted right hand movements for control than imagined right hand movements or movements of the left hand. Some studies have shown that neurons can change their tuning behavior and even reorganize with extensive practice (Ganguly and Carmena, 2010; Matsuzaka et al., 2007). As a result, the greater amount of practice with the right hand might also be related to our observed control performance results. Unfortunately, this study was not designed to evaluate the long-term effect of practice on control performance.

5 Sensory mirroring responses and mixed representations

5.1 Introduction

One of the goals of neuroscience is to understand how brain networks influence social processes and behaviors. For this reason, there has been a particular interest in the field of mirror neurons, with the majority of this field focusing on motor mirroring. Motor mirror neurons are neurons that process the performance of a motor action and the observation of someone else performing the same action similarly. These neurons have been found in both non-human primates (NHPs) as well as humans, using both electrophysiology and functional magnetic resonance imaging (fMRI) (Caggiano et al., 2009; Filimon et al., 2007; Fogassi et al., 2005; Fujii et al., 2008; Grèzes et al., 2003; Keysers et al., 2003; Mukamel et al., 2010; Rozzi et al., 2008). These studies have primarily found motor mirror neurons in brain regions involved in action planning such as the supplementary motor area (SMA), premotor cortex, and posterior parietal cortex (PPC). The prevailing theory is that the motor mirror system plays a role in simulating the inner states and intentions of other people (Caramazza et al., 2014; Keysers and Gazzola, 2006; Rizzolatti and Fabbri-Destro, 2008).

Although less well studied, sensory mirror neurons have been of increasing interest in recent years. Sensory mirror neurons are the sensory analog to motor mirror neurons, responding to the feeling of a sensation and the observation of someone else feeling the same sensation. More specifically, the sensory mirror system processes the observation of someone else feeling a sensation in the same way as it processes the actual experience of the sensation. These studies have been either electrophysiological in NHPs (Ishida et al., 2010), or using fMRI in humans (Keysers et al., 2004; Osborn and Derbyshire, 2010). Sensory mirror neurons tuned to

tactile sensations or pain have been predominantly found in the primary and secondary somatosensory cortices as well as regions around the intraparietal sulcus (IPS) in PPC.

We recorded single unit responses from the anterior intraparietal area (AIP) of a human tetraplegic volunteer (NS) participating in a brain-machine interface clinical trial, and asked how sensory mirroring responses were represented at a single neuron and population level. AIP is located within the IPS and is known to integrate visual and tactile information in the planning of grasp actions (Gentile et al., 2011; Grefkes and Fink, 2005; Klaes et al., 2015; Schaffelhofer and Scherberger, 2016) as well as in action observation (Maeda et al., 2015; Nelissen et al., 2011; Pani et al., 2014). These multisensory association properties thus make it a good candidate area for studying the sensory mirror system.

We tested neural response to “felt” sensations (when NS experienced a touch) and to “observed” sensations (when NS observed someone else getting touched). Touches (rubs) were performed on the cheek or shoulder of NS or an actor, both body parts above NS’s level of injury. We found a significant proportion of the population tuned to each of the four conditions, with populations partially overlapping, suggestive of mixed selective coding. Furthermore, the sensory mirroring component was generally confined to the same body part. Neural representations generalized better across people than body parts, suggesting a degree of functional segregation by body part. Finally, additional analyses found that the mirror responses were present regardless of whether observed sensations were presented in video compared to live action, as well as to whether NS observed the sensations with fixation or with free gaze.

5.2 Methods

5.2.1 Subject Details

Subject NS is a 59 year-old female tetraplegic 10 years post-injury, with a motor complete C3-C4 spinal lesion. She has no control or sensation below her upper arm but still retains motor control and sensations above that level. The study was approved by the California Institute of Technology, Casa Colina Centers for Rehabilitation, and the University of California, Los Angeles Internal Review Boards. We obtained informed consent after explaining the objectives of the study and the possible risks involved.

5.2.2 Experimental Setup

Experimental sessions were conducted at Casa Colina Centers for Rehabilitation. Tasks were run in a similar setup as in (Zhang et al., 2017). NS was seated in a motorized wheel chair in a lit room. For the live action tasks, an actor sat on a chair 140 centimeters in front of NS. One experimenter stood directly behind the actor while a second experimenter stood directly behind NS. The experimenters were responsible for touching NS and the actor. A 27-inch LCD monitor was positioned to the side (outside of NS's line of sight) to cue the experimenters as to what action to perform and when.

For video tasks, the videos were presented on the 27-inch LCD monitor repositioned in front of NS, occupying approximately 40 degrees of visual angle.

All stimulus presentation was controlled using the Psychophysics Toolbox (Brainard, 1997) and MATLAB. No fixation was required or enforced except in the case of the fixation vs free gaze task (below).

5.2.3 Video Recordings

The videos for the video vs live action and the fixation vs free gaze tasks were captured using a GoPro Hero 4 Silver camera mounted on a tripod. The camera was positioned such that it was at NS's normal eye position (matched in height and orientation). This positioning was done so that the recorded videos were as similar as possible to NS's perspective when viewing the actions live (in terms of distance from the scene, angle, etc.).

5.2.4 Signal Recordings

We recorded neural signals from participant NS using two 96-channel Neuroport arrays (Blackrock Microsystems model numbers 4382 and 4383) implanted in putative homologs of area AIP and Brodmann's Area 5d. Only signals from the array implanted in putative AIP were analyzed. Array placement was determined based on preoperative fMRI (Aflalo et al., 2015), with the AIP array implanted at Talairach coordinate [-36 lateral, 48 posterior, 53 superior]. Neural activity was amplified, digitized, and recorded using the Neuroport neural signal processor (NSP). The system has received FDA clearance for < 30 days acute recordings. We received FDA IDE clearance (IDE #G120096, G120287) to extend the implant duration for the purposes of the BMI clinical study.

The detection threshold for the action potentials was set at -3.5 times the root-mean-square after high pass filtering the full bandwidth signal (250 Hz cutoff). The full-bandwidth signal was sampled at 30 kHz in the Central software suite (Blackrock Microsystem). Each individual waveform was made of 48 samples (1.6 ms) with 10 samples prior to triggering and 38 samples after. Single and multiunit activity was sorted offline as in (Zhang et al., 2017).

5.2.5 Experimental Design

We used several different tasks throughout the study in order to answer specific questions about sensory mirroring responses. All tasks were similar in design and timing, differing primarily by the specific conditions involved. Unless otherwise noted, NS had her eyes open throughout the task and was asked to observe sensations "naturally", with no fixation required. Within a run, conditions were pseudorandomly interleaved, with all conditions required to be performed at least once before repeating.

In each trial, an action was cued (e.g., to touch NS's cheek) for 2.5 seconds, after which the cue disappeared. After a 3 second delay, a "Go" cue then appeared and stayed on for 4.5 seconds, signaling the experimenter to perform the previously cued action. Actions were rubs performed on the cheek/shoulder with a frequency of approximately 1 Hz (except in the case of the action tuning task, where the specific action to perform was cued). This trial structure applied to all of the tasks below.

5.2.5.1 Body part specific sensory mirroring

This task was the main data set for this study and used to study body part specific sensory mirroring responses. There were four conditions used in this task: touching NS's cheeks (felt cheek touches), touching NS's shoulders (felt shoulder touches), touching the actor's cheeks (observed cheek touches), and touching the actor's shoulders (observed shoulder touches). Touches were rubs performed bilaterally simultaneously. Cheek touches were rubs parallel to the jawline (i.e., from the cheek bone towards the chin and back again). Shoulder touches were rubs along the top of the shoulder, from near the neck to the outside of the shoulder and back again. This task was run on 6 different days, with 10 trials per condition. Eight hundred and five units were recorded in total (assuming independent recordings across days), and 756 units met unit selection criteria (see below).

5.2.5.2 Tactile receptive fields

This task was used to map out which tactile receptive fields were represented in our neural population and to identify a set of receptive fields that were relatively different from each other. Sensory mirror neurons are defined as units sharing representations across self vs other while still maintaining specificity for particular body parts. Thus, we required the body parts tested to be represented differently in order to study the sensory mirror system. For this task, there was no actor in front of NS and all actions were rubbing actions performed on NS only (i.e., only felt sensations, no observed sensations). NS had her eyes closed and wore earplugs throughout the task in order to remove any responses potentially caused by other sensory modalities (e.g., visual or audio effects), isolating the tactile sensation responses. The actions were performed on the following body parts: forehead, top of the head, back left/right of the head, left/right cheek, left/right shoulder, and the left/right side of the neck. We also included a

null condition (no sensation) as a control. Data was recorded across 2 separate days with 10 trials per condition. Two hundred and forty six units were recorded in total (assuming independent recordings across days), and 239 units met unit selection criteria (see below).

5.2.5.3 Tuning to different actions

This task was used to identify the degree to which different action types/concepts were represented compared to the strength of the mirroring responses. We did this to verify that our body part specific mirroring results were driven primarily by tuning to tactile sensations rather than action-concept tuning. We tested 16 conditions composed of the possible combinations of rubbing/tapping/pinching/pressing the cheek/shoulder on NS and the actor. Rubbing was the same action as in the body part specific mirroring task, performed at a frequency of about 1 Hz. Tapping was performed with the tips of the index and middle fingers (1 Hz). Pinching was performed with the thumb, index, and middle fingers (0.5 Hz). Pressing was performed with the index and middle fingers (0.5 Hz). This task was run on 7 different days, with 10 trials per condition. Seven hundred and seven units were recorded in total (assuming independent recordings across days), and 699 units met unit selection criteria (see below).

5.2.5.4 Video vs live action

We wanted to see the effect of having our observed sensations presented on video rather than live. To do this, we collected data with observed sensations on video on some of the same days as the main task (see body part specific sensory mirroring, above). By collecting this secondary data set on the same days as the primary data set (which had observed sensations in live action), we were able to directly compare the behavior of specific units under each of the

different presentation formats. For the video task, the timing and sequence of phases was the same as those in the main task (i.e., cue, delay, go), except instead of NS seeing a cue of what the sensation would be, she saw a still image of the first frame of the video. The video only began playing in the Go phase. The first frames of the videos were identical so that NS would not be able to learn to anticipate what action was next. For felt sensations, no video stimuli was presented. Instead, an experimenter produced them on NS live (identical to felt sensations in the main task). Videos were recorded using a GoPro camera (see Video Recordings above). This task was run on 2 different days, with 10 trials per condition. Three hundred and nine units were recorded (assuming independent recordings across days), and 252 units met unit selection criteria (see below).

5.2.5.5 Fixation vs free gaze

We wanted to assess the effect of having NS observe sensations "naturally" (i.e., with free gaze) compared to having her observe with required fixation. To do this, we added two additional sets of observed cheek/shoulder sensation conditions where NS had to fixate on one of two points, above and below the point of action contact (approximately 1.5 degrees of visual angle in either direction). All observed sensations were presented in a video format for ease of fixation point cueing (similar to the video vs live action task above). Fixation points were marked by a yellow dot on the screen, with the lack of any explicit fixation point indicating free gaze. Felt sensations were identical to those in the main task. This task was run on 2 different days, with 10 trials per condition. Two hundred and sixty four units were recorded in total (assuming independent recordings across days), and 259 units met unit selection criteria (see below).

5.2.6 Unit Selection

Within the same task, units were pooled across recording days assuming independent populations each day. To minimize low firing rate and noise effects, only units with mean firing rates greater than 0.5 Hz and an SNR > 0.5 were included in the analysis. Assessment of mean firing rates and SNR were conducted blind to the specific conditions of a trial or any analysis results.

5.2.7 Linear Analysis and Single Unit Tuning

In order to determine the tuning of individual units, we fit a linear model to each unit's firing rates during a "Go" and baseline time window. The Go time window was 1 to 4 seconds after sensation onset to isolate the stimulus response, mitigating possible cases where the experimenters may have started the touch late or ended the touch early. The baseline time window was 1.5 to 2.5 seconds after delay phase onset in order to capture the neural response in the absence of any stimulus (including visual differences from the experimenters reading the onscreen cue and possible early touches by the experimenters). We fit the unit's firing rate as a function of condition indicator variables as in (Zhang et al., 2017):

$$FR = \sum_{c} \beta_c X_c + \beta_0$$

where FR is the firing rate, X_c is the vector indicator variable for condition c, β_c is the estimated scalar weighting coefficient for condition c, and β_0 is a constant offset term. In this model, the beta coefficients represent the expected firing rate changes from baseline for each condition. For each condition, the indicator variable is a vector of binary values where each element is 1 if the corresponding data point at that index is of the same condition type, and 0 if the data point

is of a different condition type. All baseline samples were also assigned a 0, effectively pooling together baseline data independent of condition. A unit was considered tuned to a condition if the t-statistic for the beta coefficient associated with the condition was significant ($p < 0.05$, FDR corrected).

We tested whether the fraction of units tuned to each condition was significantly different using a non-parametric ANOVA (Kruskal-Wallis) on the significance values of the units for each condition (as described above). Comparisons were performed for each pair of conditions.

5.2.8 Degree of Specificity

We wanted to assess the degree to which the populations tuned to each of the conditions were overlapping/specific. We did this by computing how specific each unit was to each of the conditions, comparing conditions in a pairwise fashion (e.g., how specific a unit was to felt cheek sensations vs observed cheek sensations, felt cheek sensations vs felt shoulder sensations, etc.). We computed the degree of specificity by taking the difference of the absolute values of the beta coefficients (associated with the two conditions being compared) normalized by the sum of the absolute values of the beta coefficients. For example, to compute the degree of specificity to felt and observed cheek sensations, the equation would be:

$$Degree\ of\ Specificity = \frac{|\beta_{FeelCheek}| - |\beta_{ObsCheek}|}{|\beta_{FeelCheek}| + |\beta_{ObsCheek}|}$$

where β is the beta coefficients from the linear model fit above for the associated movement condition. This degree of specificity analysis was done separately for each pairwise combination of conditions. We only included units tuned to at least one of the conditions being compared (p

< 0.05, FDR corrected). For each distribution, we also performed a two-sided sign test to determine whether the medians of the distributions were significantly different from 0 (i.e., if they were biased to one variable over another).

5.2.9 Confusion Matrix

We wanted to verify that it was possible to decode each of the tested conditions from each other, despite the populations underlying the neural representations being highly overlapping. To do this, we used a confusion matrix analysis, computing the frequency (on a trial-by-trial basis) with which each condition was misclassified by a classifier.

For the classifier, we used a linear classifier (linear discriminant analysis, all classes assumed to have the same diagonal covariance matrix). For features, the classifier used the mean firing rates during the Go phase subtracted by the corresponding mean baseline firing rates, computed on a trial-by-trial basis. The time windows were the same as the ones used for the linear models above. Performance was calculated using leave-one-out cross-validation. The confusion matrices were computed for each day separately, and then averaged together for Figure 5.1D.

5.2.10 ANOVA Analysis

Besides identifying units by their tuning to each individual condition, we also wanted to categorize units by their tuning to the body part, person, and the interaction between the two. On a per unit basis, we performed a 2-by-2 ANOVA on the unit's firing rate modulation from baseline (Go phase activity subtracted by mean baseline activity, using the same windows as in

the linear analysis above). We grouped the data by the person variable (felt vs observed sensations) and the body part variable (cheek vs shoulder) associated with each condition, testing the effects of each as well as the interaction effects (i.e., 3 terms: person, body part, and the interaction of person and body part). A unit was considered to have a significant effect of a variable if the associated p-value was significant ($p < 0.05$, FDR corrected). Units were considered body part specific mirror neurons if they showed a significant main effect of body part and no significant interaction effect.

5.2.11 Principal Component Analysis (PCA)

We wanted to identify the most natural modes of our network to see what variables our neural population encoded the most strongly. We performed a principal component analysis (PCA) to reduce our high dimensional firing rate data into 10 principal components (only the first four are shown in Figure 5.2C-F). We performed the PCA on the trial-by-trial firing rate data. Thus, the PCA was applied to a matrix of (Time Samples Per Trial * Conditions * Trials Per Condition) x (Number of units), with the number of units being the variables. Data was pooled across days assuming independent populations across recording sessions. The entire trial time window was included, from 0.5 seconds before the trial started to the end of the trial (4.5 seconds after Go onset), resulting in a total time window of 10.5 seconds. Neural data was sampled with 50 ms time bins and smoothed using a minimum jerk smoothing kernel (500 ms smoothing window). Finally, in generating Figure 5.2C-F, we computed the projections of each trial's neural data to each of the principal components and then grouped the trials by their condition type.

5.2.12 Correlation of Neural Responses

We wanted to examine how the different conditions were represented relative to each other and specifically which conditions had more similar neural representations. Using correlation as a measure of similarity, we compared the conditions in a pairwise fashion. We used correlation over other distance measures (such as Euclidean or Mahalanobis distance) because correlation gives a normalized value of the similarity that is agnostic to gross baseline changes across the entire population.

We quantified the neural representations as the vector of normalized beta coefficients (beta coefficients from the linear model fits above divided by their 95% confidence intervals, one element per unit). The normalized beta coefficients are thus a trial average measure of each unit's firing rate modulation from baseline weighted by its trial-to-trial variability. Only units significantly tuned to at least one of the conditions ($p < 0.05$, FDR corrected) were included in the correlation analysis.

5.2.13 Hierarchical Clustering or Neural Responses

The correlation analyses above suggested a representational structure where some conditions were more similar to each other than to others. To summarize the structure of the representations and how representations were clustered together, we performed a hierarchical clustering analysis (agglomerative hierarchical cluster tree). We used the built-in MATLAB 2017a linkage and dendrogram functions with weighted average correlation as the measure (McKenzie et al., 2014; Zhang et al., 2017).

5.2.14 Cross Decoding

We wanted to test the degree to which representations of one variable generalized across the different levels of the other variable. In Figure 5.3D, for example, we tested how well representations of person generalized across the different body parts. We used a cross-decoding analysis similar to the one used in (Zhang et al., 2017). We trained a classifier on only one level of a variable to classify the levels of the other variable. We then computed both its cross-validated performance (on the training data) and its generalization performance when applied to the untrained data (from the other level of the variable). This was done in both directions (i.e., training on the other level of a variable). In the example of Figure 5.3D, we trained a classifier on cheek data to classify between felt and observed sensations (blue bars). We then computed the cross-validated performance (left blue bar), and the generalization performance (right blue bar). We then repeated this analysis in the opposite direction, training on shoulder data and testing on cheek data.

For the classifier, we used a linear classifier (linear discriminant analysis, diagonal linear covariance matrix). For features, we used the average firing rates during the Go window subtracted by their corresponding baseline window activity (taken on a trial-by-trial basis, with time windows the same as in the linear analysis above). Neural data was pooled across days, assuming independent neural populations across recording days.

Cross-validation performance was determined by leave-one-out cross-validation. Performance measures were simply classification error (i.e., fraction of trials for which the predicted label matched the true label). Significance was determined by a permutation test (shuffling the true

labels to create a null distribution, 1000 shuffles), and comparing the true value to the null distribution. A value was considered significant if it was greater than 95% of the values in the shuffle distribution values (i.e., $p < 0.05$).

5.2.15 Number of Units Invariant Analysis

We examined the likelihood of units being invariant to multiple body parts in our neural population, even beyond the cheek and shoulder tested in our main data set (e.g., neck, forehead, back of the head, etc.). We looked at the fraction of the population invariant to different numbers of body parts (N), using just the data from the tactile receptive field task. We computed the fractions for all of the possible combinations of the N body parts. For example, for N = 1, we looked at the percent of units tuned to each of the 10 tested body parts (i.e., 10 values). For N = 2, we looked at the percent of units tuned to each of the possible pairs of body parts (10*9/2 = 45 values). A unit was considered tuned to all N body parts if the significance values from the linear model fits were significant for all of the body parts involved ($p < 0.05$, FDR corrected). N varied from 1 to 10 (10 total body parts). For each value of N, we averaged the percent of units tuned to generate the curve in Figure 5.4C.

5.3 Results

5.3.1 Sensory Mirroring

To study sensory mirroring, we recorded from AIP of a female C3/C4 tetraplegic participant (NS). We compared the neural responses to felt/observed sensations of the cheek and shoulder. “Felt sensations” refers to touches felt by NS while “observed sensations” refers to NS observing

touch actions performed on another person. Sensory mirroring data was collected using a delayed action paradigm (Figure 5.1A). Two experimenters and one actor were also involved during data collection. The actor was seated across from NS while one experimenter stood behind NS and one experimenter stood behind the actor. At the start of each trial, experimenters were cued for 2.5 seconds on *who* to touch (NS vs the actor) and *what* to touch (the cheek or shoulder). After an additional delay of 3 seconds, a "Go" cue was displayed for 4.5 seconds, cueing the experimenters to perform the action. NS's eyes were open throughout the task, and was instructed to observe actions "naturally". We were careful to ensure that NS could not see the cues and thus could not anticipate the actions. Actions were performed on the cheek/shoulder based on a preliminary receptive field test (see below), with both body parts above NS's level of injury and thus still sensitive. In addition, we performed all actions live and with no fixation enforced (more on that below).

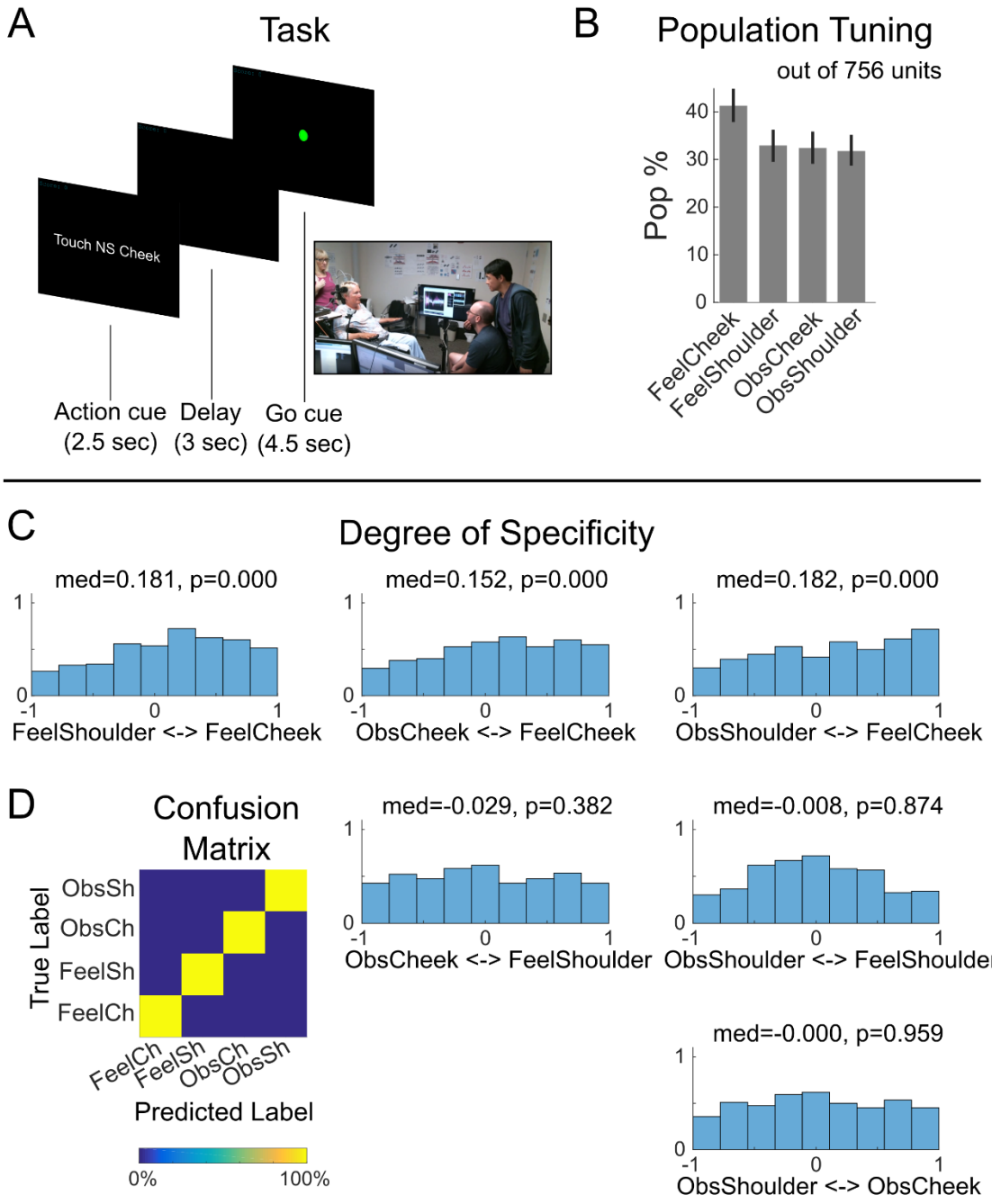

Figure 5.1. Overlapping Populations Tuned to Felt and Observed Sensations

(A) General data collection paradigm showing an example action cue. The green dot is the "Go" cue. (B) Percent of units tuned to each of the four tested conditions (bootstrap 95% CI, $p < 0.05$, FDR corrected). (C) Distribution of the degree of specificity for individual neurons to all the different possible pairwise combinations of conditions, plotted as the estimated probability density function. For each distribution, the median and the probability of the median being different from 0 (two-sided sign test) are shown. (D) Confusion matrix showing the percent of the time each condition was classified as each other condition. The matrix shown is the average of the confusion matrices computed on each day individually. (Feel = felt, Obs = observed; Ch = cheek, Sh = shoulder).

We first examined the tuning in the neural population to each of the four tested conditions: felt cheek touches, felt shoulder touches, observed cheek touches, and observed shoulder touches. A significant percent of the population was tuned to each, with a greater fraction tuned to felt cheek touches (Figure 5.1B, $p < 0.001$ for each pairwise comparison of felt check touch to the other conditions using a Kruskal-Wallis test on the significance values of each condition).

We next asked to what degree the populations tuned to felt and observed sensations were overlapping using a degree of specificity analysis. The potential mirror neurons would exist at this overlap, being nonspecific for felt vs observed sensations. For each pair of conditions, we computed the degree of specificity of each unit as the difference in the absolute values of the beta coefficients normalized by the sum of their absolute values. This results in a value between -1 and 1, with 0 indicating nonspecific tuning and 1 and -1 indicating highly specific tuning. The distributions of these values for each pairwise comparison are shown in Figure 5.1C. The neural populations tuned to felt and observed sensations of the same body part were indeed overlapping, suggesting the presence of sensory mirror neurons in the population. However, neural populations tuned to other pairs of conditions were similarly overlapping.

Despite this high degree of overlap, we were still able to separate all four conditions from each other. We assessed the cross-validated classification performance of a linear classifier trained to decode the four conditions from each other and computed the confusion matrix for each day. The average of these confusion matrices is shown in Figure 5.1D.

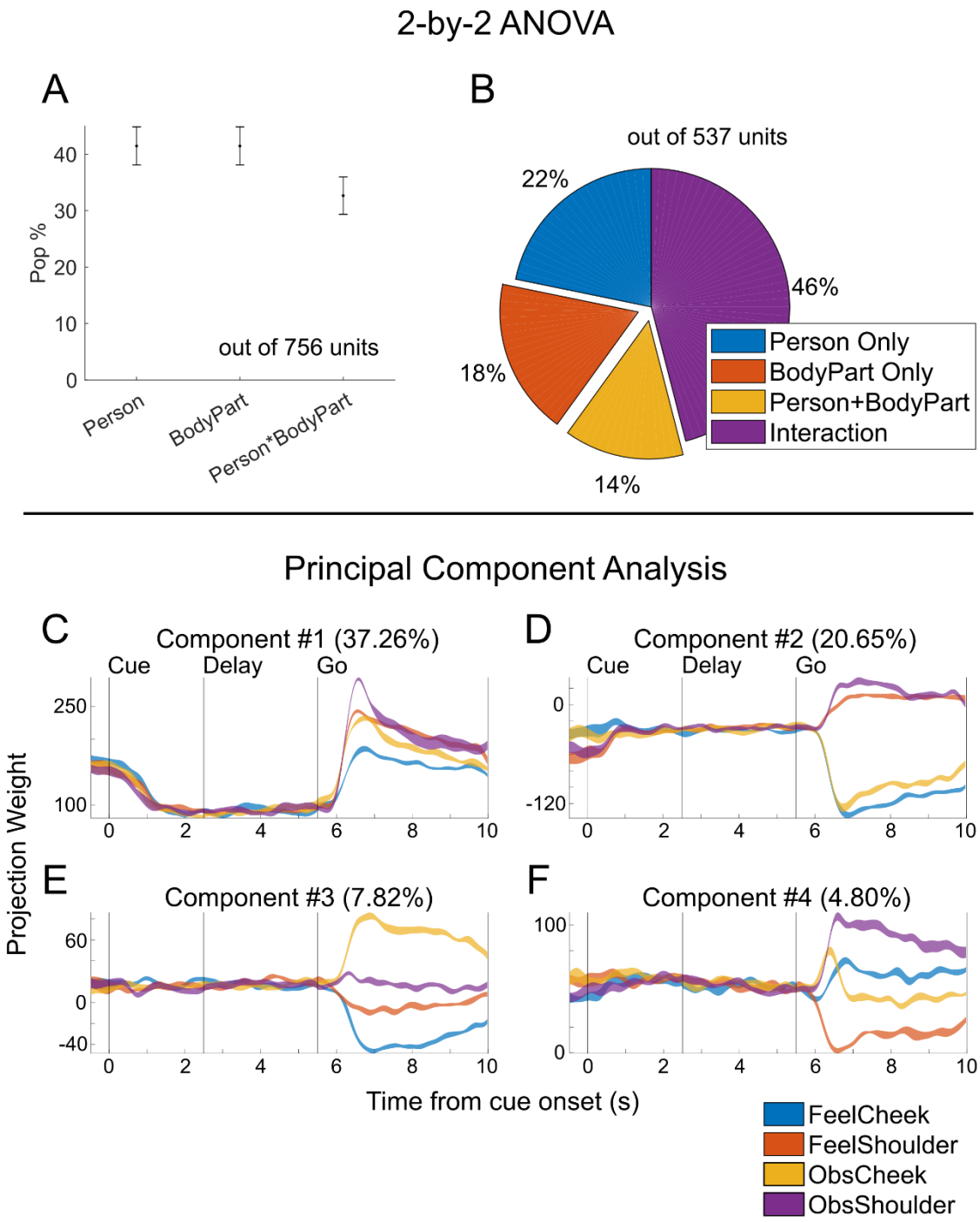

Figure 5.2. Sensory Mirroring Responses

(A) Percent of the population showing a significant effect of each term in the 2-by-2 ANOVA (bootstrap 95% CI, $p < 0.05$, FDR corrected). (B) Pie chart showing distribution of units by their tuning type as determined by a 2-by-2 ANOVA (e.g., main effect of person only, main effect of body part only, etc.). Only units with tuning to at least one of the terms in the ANOVA were included. The exploded sections (red and yellow) represent those that would fall into our definition of "body part specific mirror units". (C) First four components from a principal component analysis of the neural data plotted against time. Trial-by-trial responses were projected onto each of the four conditions and their trial average and standard error plotted. The percent of the variance accounted for by each component is also displayed.

We next asked if the sensory mirroring responses in the population were significant relative to the other encoded variables. First looking at the single unit level, we performed a 2-by-2 ANOVA on each unit independently, testing the effects of the person variable (i.e., whether the sensation was felt or observed; who was touched), the body part variable (i.e., whether the cheek or the shoulder was touched), and their interaction. A neuron was considered body part specific if it showed a main effect of body part and no interaction effect, as this would mean it responded differently to different body parts ("body part specific") and in a similar way regardless of who was touched ("mirroring"). Units with a main effect of both body part and person were still considered mirroring, as they would be coding the body parts similarly, albeit with a mean offset based on the person touched. Looking at the percent of the population showing a significant effect of each term, the proportion of units with a main effect of body part was comparable to the proportion of units with a main effect of person ($p < 0.05$, FDR corrected, Figure 5.2A). Furthermore, the fraction of units with a main effect of body part alone was comparable to the fraction of units with a main effect of person alone (Figure 5.2B). Based on our definition of sensory mirror units, a significant fraction of the population was sensory mirroring (Figure 5.2B, red and yellow sections).

We also directly visualized this mirroring component at the population level, performing a principal component analysis (PCA) on the firing rate data of the population as a whole. The first four principal components are shown in Figure 5.2C-F. The largest component (37.26% of the variance, Figure 5.2C) was consistent with a general sensation/action onset component. The second component (20.65% of the variance, Figure 5.2D) modulated similarly for felt and observed cheek touches (blue and yellow traces), and another way for felt and observed shoulder touches (red and purple traces). This component behaved as a body part specific

mirroring component, treating sensations on the same body part essentially identically, regardless of who was touched. The fact that this mirroring component occupied so much of the variance in the population is consistent with a dominant mirroring component.

Interestingly, none of the other principal components (neither the two shown in Figure 5.2EF nor any of the others) behaved as a "person-specific" counterpart to the mirror component, i.e., treating sensations on the same person identically regardless of what was touched. This suggests that the body part variable is coded specially at a population level compared to the person variable.

To test this directly, we performed a correlation analysis on the neural responses to each of the four conditions. We found that the representations of felt and observed sensations of the same body part were the most similar to each other (Figure 5.3A). This was further illustrated by a hierarchical clustering analysis performed on the normalized beta coefficients, using correlation as a distance measure (Figure 5.3B). In other words, representations that differ by person are more similar to each other than representations that differ by body parts. The differences in the similarity suggests that the body part and person variables are not randomly mixed. Instead, the variables are partially mixed, with representations of person organized (i.e., functionally segregated) by body part.

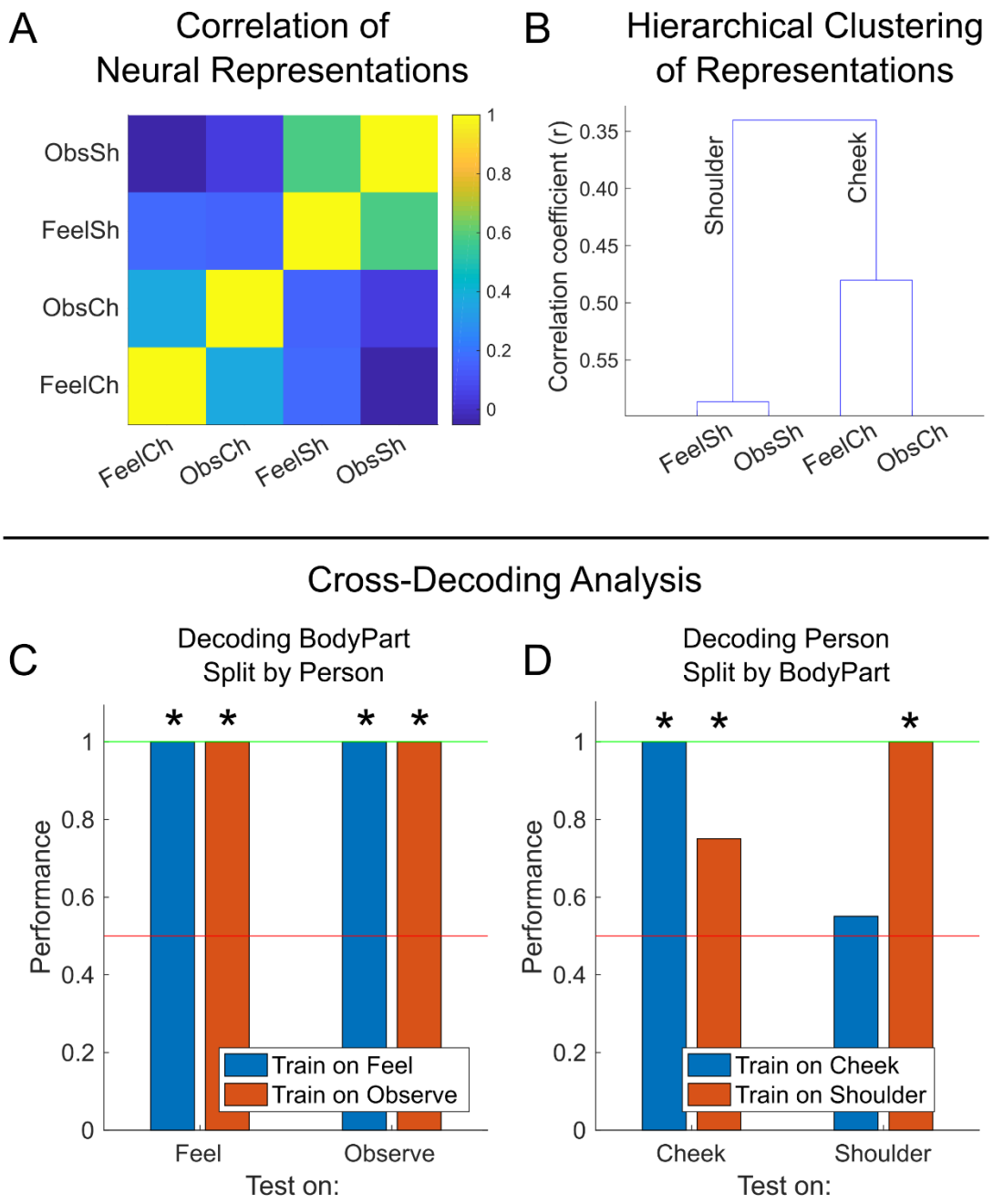

Figure 5.3. Representations Organized by Body Part

(A) Correlation matrix showing the similarity between the neural representations of the four conditions. (B) Dendrogram summarizing the hierarchical clustering of the neural representations of the four conditions, using correlation as a measure. The primary splitting variable of the major branches is shown. Representations were mainly split by body part. (C) Cross-decoding analysis showing the generalization performance of person representations across person. In blue is the performance of a classifier trained on "feel" data to classify between cheek/shoulder applied to feel data (cross-validated performance, left blue bar), and "observe" data (generalization performance, right blue bar). In red is the performance of a classifier trained on observe data applied to observe data (cross-validated performance, right red bar) and feel data (generalization performance, left red bar). A red and green line marks chance and maximum performance, respectively. Performance significantly above chance is also marked (permutation test, $p < 0.05$, uncorrected). (D) Similar to (C) but with classifiers trained on only cheek/shoulder data to classify between felt/observed sensations (i.e., generalization of person representations across body parts).

The greater similarity between representations differing by people compared to representations differing by body part suggests that representations may generalize better across different people than different body parts. We directly tested for this possibility by performing two cross-decoding analyses similar to the ones used in that study. For the first cross-decoding analysis (Figure 5.3C), we tested the degree to which representations of body part generalized between the person being touched (i.e., felt vs observed sensations). We trained a linear classifier to classify the body part touched based on only felt sensations and then evaluated the classifier's performance on felt sensations (cross-validation performance, left blue bar) as well as observed sensations (generalization performance, right blue bar). We also trained a second classifier on the reverse direction, training on only observed sensations and testing on observed sensations (cross-validation performance, right red bar) and felt sensations data (generalization performance, left red bar). Both classifiers had cross-validation and generalization performance significantly above chance ($p < 0.05$, one-sided permutation test). This suggests that the representations of body parts do generalize across felt and observed touches.

For the second cross-decoding analysis (Figure 5.3D), we tested how well the representations of the person being touched (i.e., whether the sensation was felt or observed) generalized across different body parts. In this case, we found that the representations generalized somewhat across different body parts (at least from shoulder to cheek), albeit imperfectly. Altogether, the cross-decoding analyses show that the representations in our neural population generalize better across different people than across different body parts, further showing that the representation of body part is special in our neural population.

These cross-decoding results are consistent with a population-level sensory mirroring response. Single unit sensory mirror neurons are defined as units that process observed sensations similarly to felt sensations when matched by body part. Analogously, a population-level sensory mirroring response can be defined by the encoding of observed sensations similarly to felt sensations when matched by body part. This similarity would make the structure of body part representations similar between felt sensations and observed sensations, allowing body representations to generalize from one to the other. In other words, the result in Figure 5.3C illustrates a defining property of sensory mirroring: that the representations of felt sensations generalize to observed sensations.

5.3.2 Tactile Receptive Fields

Before collecting any sensory mirroring data, we first wanted to identify tactile receptive fields in NS that were relatively orthogonally represented. Sensory mirroring requires similar responses between observed and felt sensations specific to body parts. Without this specificity, any shared responses between felt and observed sensations could just be responses to an action occurring, the concept of the action, or some other shared feature. To find specificity, we needed the body parts to be differently represented. Thus, we needed to find receptive fields as different as possible in representation in order to observe a body part specific mirroring effect.

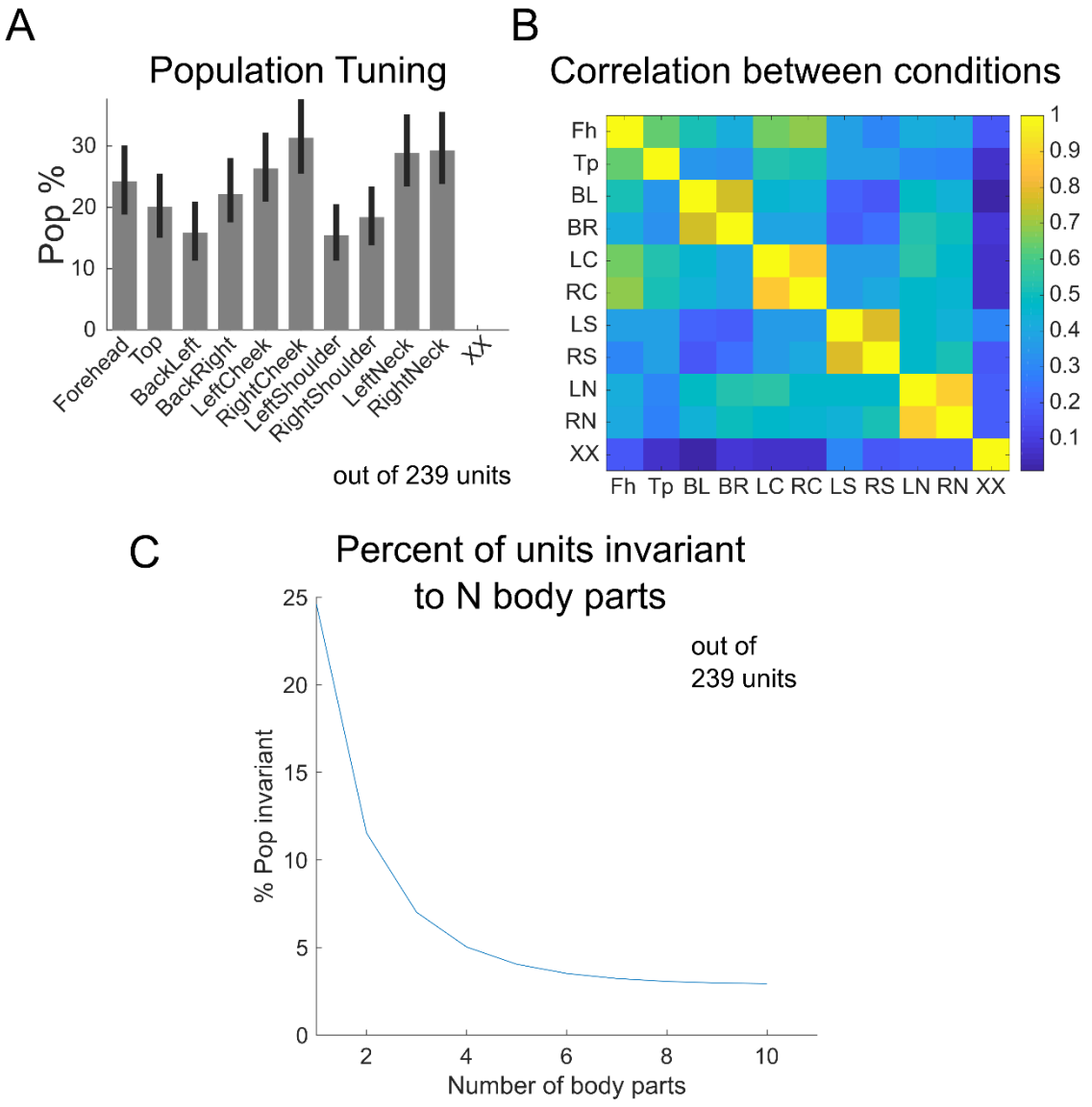

Figure 5.4. Tactile Receptive Fields

(A) Percent of the population significantly tuned to each of the tactile receptive fields stimulated (bootstrap 95% CI, $p < 0.05$, FDR corrected). (B) Correlation matrix showing similarity between the neural representations of each of the tactile receptive fields. (Fh = forehead, Tp = top of head, BL = back left of head, BR = back right of head, LC = left cheek, RC = right cheek, LS = left shoulder, RS = right shoulder, LN = left side of neck, RN = right side of neck, XX = no stimulation). (C) Plot showing the average percent of units tuned and invariant to N different body parts, as N increases from 1 to 10 ($p < 0.05$, FDR corrected).

To do this, we used a delayed action paradigm where the experimenter touched NS on different parts of the body above the level of injury while her eyes were closed. Touches were rubbing motions of approximately 4 Hz in frequency and for a duration of approximately 4 seconds. The different receptive fields tested were: forehead, top of the head, back left/right of the head,

left/right cheek, left/right shoulder, and the left/right side of the neck. We also included a null (no sensation) condition as a control (see Methods for more details).

Looking at the percent of the population tuned to each of the receptive fields, we found that all tested receptive fields were represented in our neural population to varying degrees (Figure 5.4A). As expected, there was no significant tuning to the null control condition ("XX").

Using correlation as a measure, we next looked at the similarity in the neural representations (Figure 5.4B). We computed the correlation by correlating the normalized beta coefficients from the linear model fits for each condition (beta coefficients normalized by their 95% confidence intervals) against each other (see Methods for details). We found that body parts were relatively similar in representation bilaterally (e.g., left and right shoulder were much more similar to each other than anything else). Furthermore, the representations of all the tested tactile sensations were most different from the null condition. Based on the similarity between the representations, we decided to use the cheek and shoulder as the main body parts in our study. These body parts were not only relatively different in representation but also easily distinguishable visually. It was important to choose sensations visually distinguishable so that the observe conditions could be well differentiated from each other as well.

In Figure 5.2B, we found a significant fraction of units that were specific to who was being touched but were invariant to the body part touched, essentially "person-specific". However, it is possible that these "person-specific" units just happen to be units that are invariant to the

tested body parts (cheek and shoulder), and not necessarily all body parts. We tested for this possibility by looking at the fraction of units in our neural population that were tuned and invariant to all the possible combinations of the body parts tested in the receptive field task (10 body parts). As more body parts were introduced to the comparison, the number of units tuned to all of them decreased (Figure 5.4C). Only units invariant to all body parts could have the potential to be person-specific. This suggests that many of the "person-specific" units identified in Figure 5.2B, while invariant to cheek and shoulder, are not truly person-specific and would actually show body part-specific effects if more body parts were tested in the task.

5.3.3 Tuning to Action Concepts

Past studies have found the region we are recording from (AIP) to be associated with action observation and even action understanding, with neurons responding differentially when observing different grasps (Fogassi et al., 2005; Nelissen et al., 2011; Pani et al., 2014; Wurm and Lingnau, 2015). One might argue that the sensory mirroring responses are actually responses coding the concept of specific touch actions. In the above analyses, we found responses matched by body part that were consistent between different people, interpreting them as sensory mirroring results. However, tuning to different action concepts (e.g., the concepts "cheek touches" vs "shoulder touches", independent of whose cheek/shoulder) would also explain these results.

We differentiated between these two possibilities by running a task similar to the primary task but with additional actions interleaved. Besides rubbing the cheek/shoulder, the experimenters also pinched, pressed, and tapped the cheek and shoulder on both NS and the actor (4 actions,

16 conditions total). All 16 conditions were represented in the population (Figure 5.5A). Correlating the representations against each other, we found a significant block diagonal structure (Figure 5.5B). Most notably, conditions involving the same body part and person (i.e., differing only by action) were the most similar to each other; conditions differing only by body part or person were much less correlated in contrast (Figure 5.5C). This suggests that, although the different actions are indeed represented in our neural population, the differences between representations are dominated by body part and person tuning. In other words, the effects of action concept tuning are relatively minor compared to the effects of body part and person.

We also performed a hierarchical clustering analysis to visualize this. Using correlation as a measure, we clustered the neural responses to each of the 16 conditions (Figure 5.5D). Representations were primarily clustered by body part, followed by person, and lastly by the specific action.

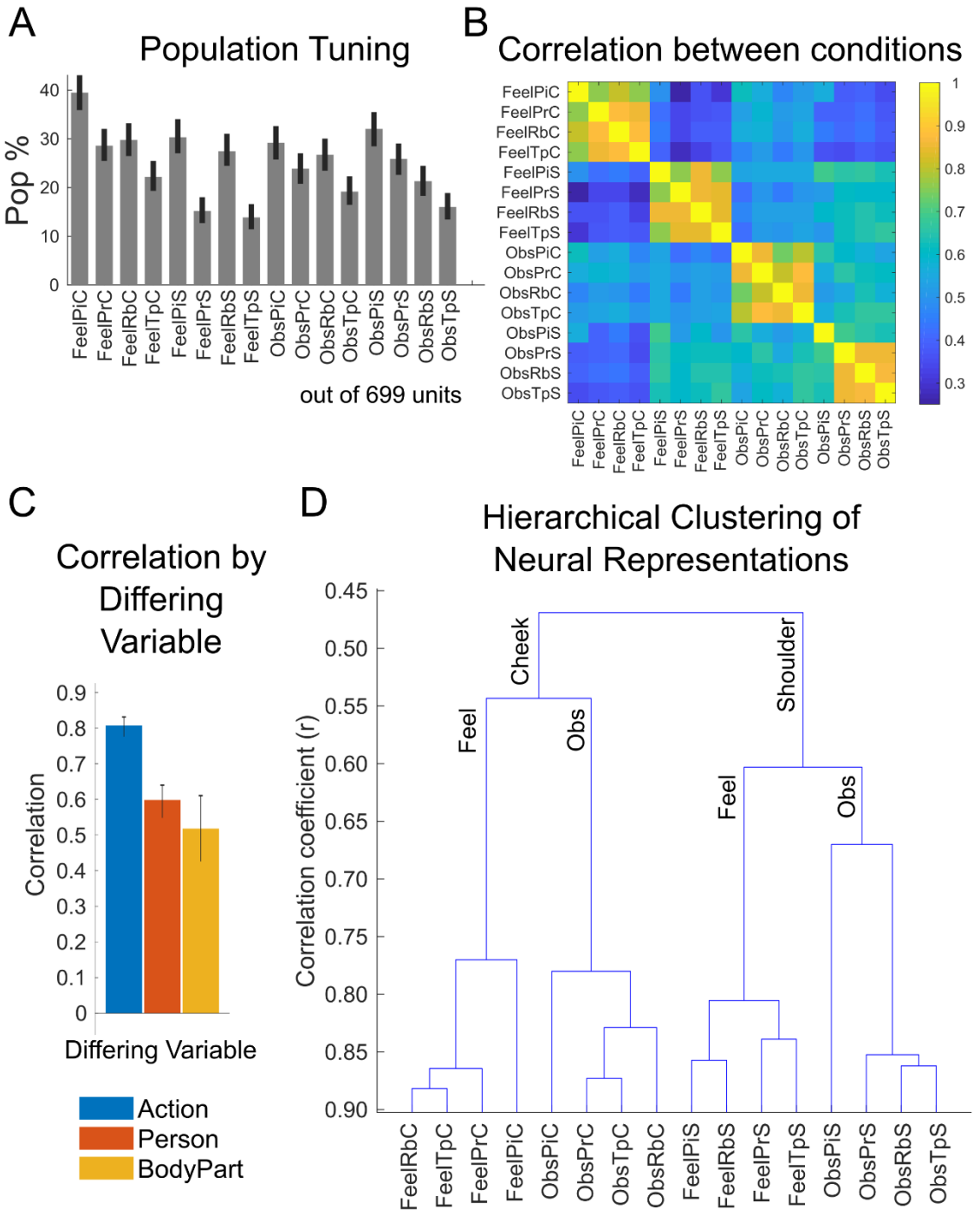

Figure 5.5. Effect of Action Observation

(A) Percent of the population significantly tuned to each of the specific action/person/body part combinations (bootstrap 95% CI, $p < 0.05$, FDR corrected). (B) Correlation matrix showing the degree of similarity between each of the 16 tested conditions. (C) Correlation between pairs of conditions (i.e., values from (B)) grouped by the differing variable. The average correlations within each group and the standard errors are displayed. For example, the blue bar represents the average correlation of all pairs of actions from (B) where both were performed on the same person and body part, but differed only by whether they were a pinch, press, rub, or tap. (D) Dendrogram summarizing the hierarchical clustering structure of the neural representations. The variable primarily splitting each major branch is labeled. Representations were first split by body part, and then by person. (Feel = felt, Obs = observed; Pi = pinch, Pr = press, Rb = rub, Tp = tap; C = cheek, S = shoulder).

5.3.4 Effects of Video vs Live Action

In our task paradigm, NS observed actions performed live, rather than actions performed in video. To test for the effects of presentation format, we ran a secondary experimental task during several of the same recording sessions as the primary task. In this task, all observed sensations were presented in video on an LCD screen placed in front of NS. The videos were recorded using a GoPro camera placed near NS's head at eye level, in order to most similarly capture NS's perspective when viewing the actions live. Felt sensations were the same as in the primary task, with an experimenter producing them live.

The main question we wanted to answer was whether our primary result (i.e., body part specific mirroring) would be robust to changes in the presentation format for observed actions. We first verified that the observed sensations were represented in both live action and video. Looking at the single unit tuning, we found that each of the six conditions (felt/observed-live/observed-video cheek/shoulder sensations) were represented by a significant fraction of the population, with more for the live action than the video format (Figure 5.6A).

The sensory mirroring result can be succinctly represented as the generalization of body part representations from felt sensations to observed sensations (Figure 5.3C), demonstrating that the neural population processes observed sensations similarly to experienced sensations. Thus, we directly tested for this mirroring result by performing a cross-decoding analysis to see how well representations of body part generalized across NS's felt sensations, observed sensations in live action, and observed sensations in video. While the representations did not generalize as well from observed sensations to felt sensations (Figure 5.6B, left red and yellow bars), the felt

sensations did generalize well in the opposite direction (Figure 5.6B, blue bars). This suggests that, although representations across the population are somewhat affected by whether the observed sensations are in live action or video, the overall mirror response behavior (i.e., generalization of representations from felt sensations to observed sensations) still exists regardless of format.

As opposed to looking at the population as a whole, we also looked at the degree to which the specific sensory mirror neurons (i.e., body part specific units) were differentially tuned to video vs live action. We first performed a 2-by-2 ANOVA to identify the sensory mirror units (similar to the method in Figure 5.2B, testing the effects of person and body part) (Figure 5.6C, $p < 0.05$, FDR corrected). We performed this analysis with respect to only the felt and observed live conditions to be more consistent with Figure 5.2B in methodology and to identify a more similar population of sensory mirror units (i.e., units showing a main effect of body part but no interaction effect during the live action case). To test how well the representations of body part generalized across the formats and felt sensations for these mirror neurons specifically, we performed another cross-decoding analysis, similar to the one in Figure 5.6B. Using only the body part specific neurons (Figure 5.6C, red and yellow sections), we found that the representations were indeed preserved (Figure 5.6D). This suggests that the sensory mirroring responses exist in both live action and video formats, albeit with relatively poorer generalization between felt and observed touches in video.

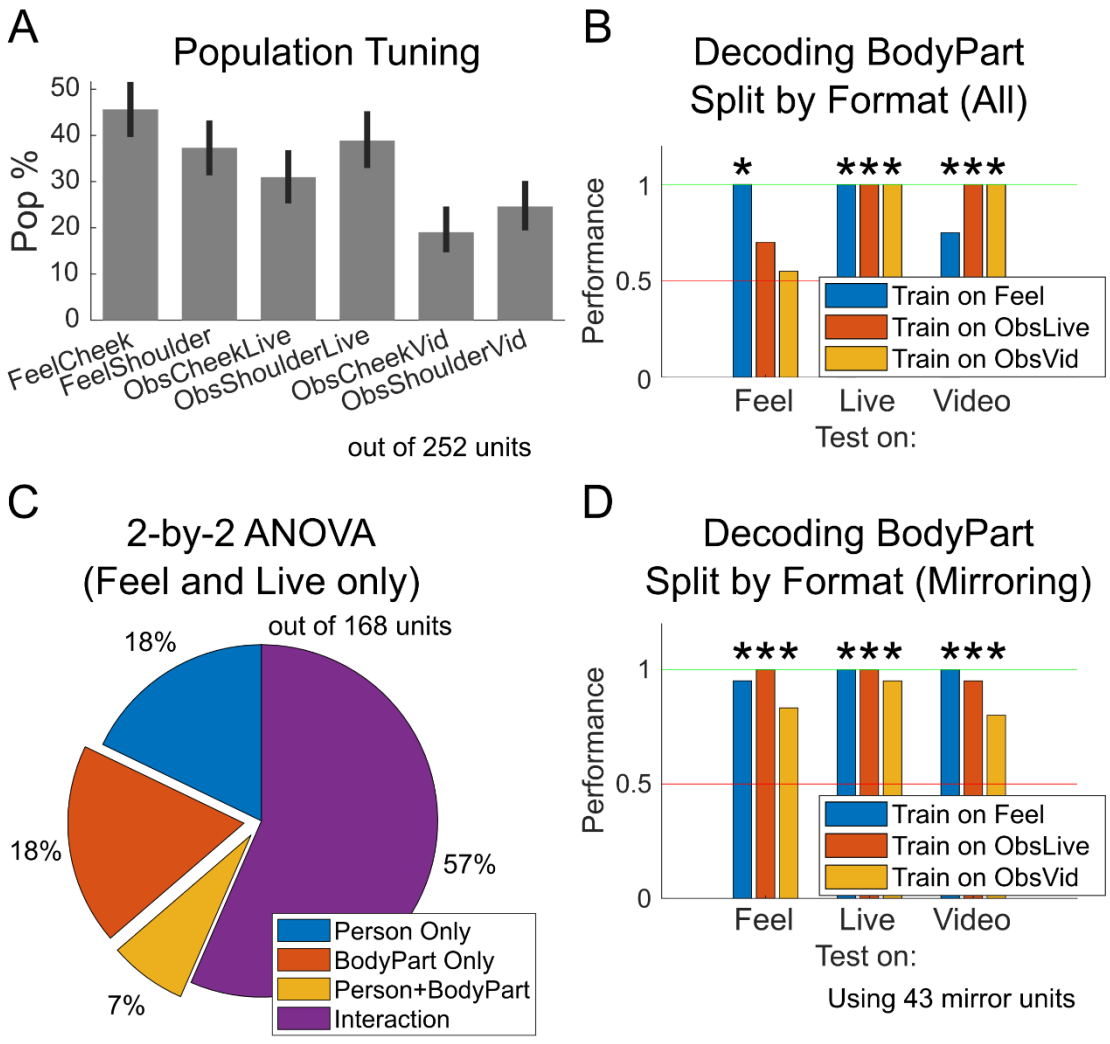

Figure 5.6. Effect of Presentation Format (Live Action vs Video)

(A) Percent of the population significantly tuned to each of the feel conditions and live action/video conditions (bootstrap 95% CI, $p < 0.05$, FDR corrected). (B) Cross-decode analysis performed on the entire neural population, similar to the one in Figure 5.3C, testing generalization of body part representations across felt sensations and live action/video. (C) Pie chart similar to the one in Figure 5.2B, showing the unit tuning types as determined by a 2-by-2 ANOVA. Only felt and live action conditions were included in the analysis and only units with tuning to at least one of the terms in the ANOVA were included in the pie chart. (D) Cross-decode analysis, similar to (B), but performed on only the body part specific mirror units (i.e., the ones in the yellow/red sections in (C)). (Feel = felt, Obs = observed; Live = live action, Vid = video).

5.3.5 Effects of Fixation vs Free Gaze

In the main task, we instructed NS to observe actions "naturally", using free gaze as opposed to requiring eye fixation. To test the effects of fixation condition, we ran an additional experiment

where NS was sometimes asked to fixate and other times allowed to free gaze while observing sensations. There were two fixation points for each body part, one above and one below the point of contact (about 1.5 degrees of visual angle above/below). All observed sensations were also presented in video, so that it was easier to cue fixation points and control the depth of field.

We did a set of analyses similar to those used to assess the effects of video vs live action above (Figure 5.6). We first confirmed that our neural population was tuned to each of the fixation conditions, and indeed, a significant fraction of units were tuned to each of the fixation points, with a greater proportion tuned to free gaze than for either of the fixation conditions (Figure 5.7A). To check that the sensory mirroring was present across the fixation conditions, we performed a cross-decoding analysis (using all the units in the population), assessing how well representations of body part generalized across NS's felt sensations, observed sensations with free gaze, and observed sensations with each of the fixation points (Figure 5.7B). Similar to the result from Figure 6B, while the representations did not generalize well from observed sensations (both free gaze and fixation) to felt sensations (Figure 5.7B, left red, yellow, and purple bars), the felt sensations did generalize in the other direction (Figure 5.7B, blue bars). This suggests that, generally, the mirror responses in the population (i.e., generalization from felt to observed) are not significantly affected by fixation condition.

Besides looking at how representations are affect by fixation across the entire population, we also focused specifically on the sensory mirror neurons. We once again used a 2-by-2 ANOVA to identify the body part specific neurons, i.e., those having a main effect of body part and no interaction between body part and person (Figure 5.7C, $p < 0.05$, FDR corrected). We performed

this analysis only on the felt and free gaze conditions so as to resemble more closely the original 2-by-2 ANOVA results in Figure 2B (which used felt and free gaze live action conditions) and identify a more similar population of sensory mirror units. We then performed another cross-decoding analysis, looking only at these sensory mirror units (Figure 5.7C, red and yellow sections), to assess how much fixation condition affects the mirror neurons' body part representations. The body part representations (at least the representations learned from felt sensations) were largely preserved (Figure 5.7D, particularly the blue bars). Altogether, these results suggest that fixation condition does not have a significant effect at a representational level on the sensory mirroring responses.

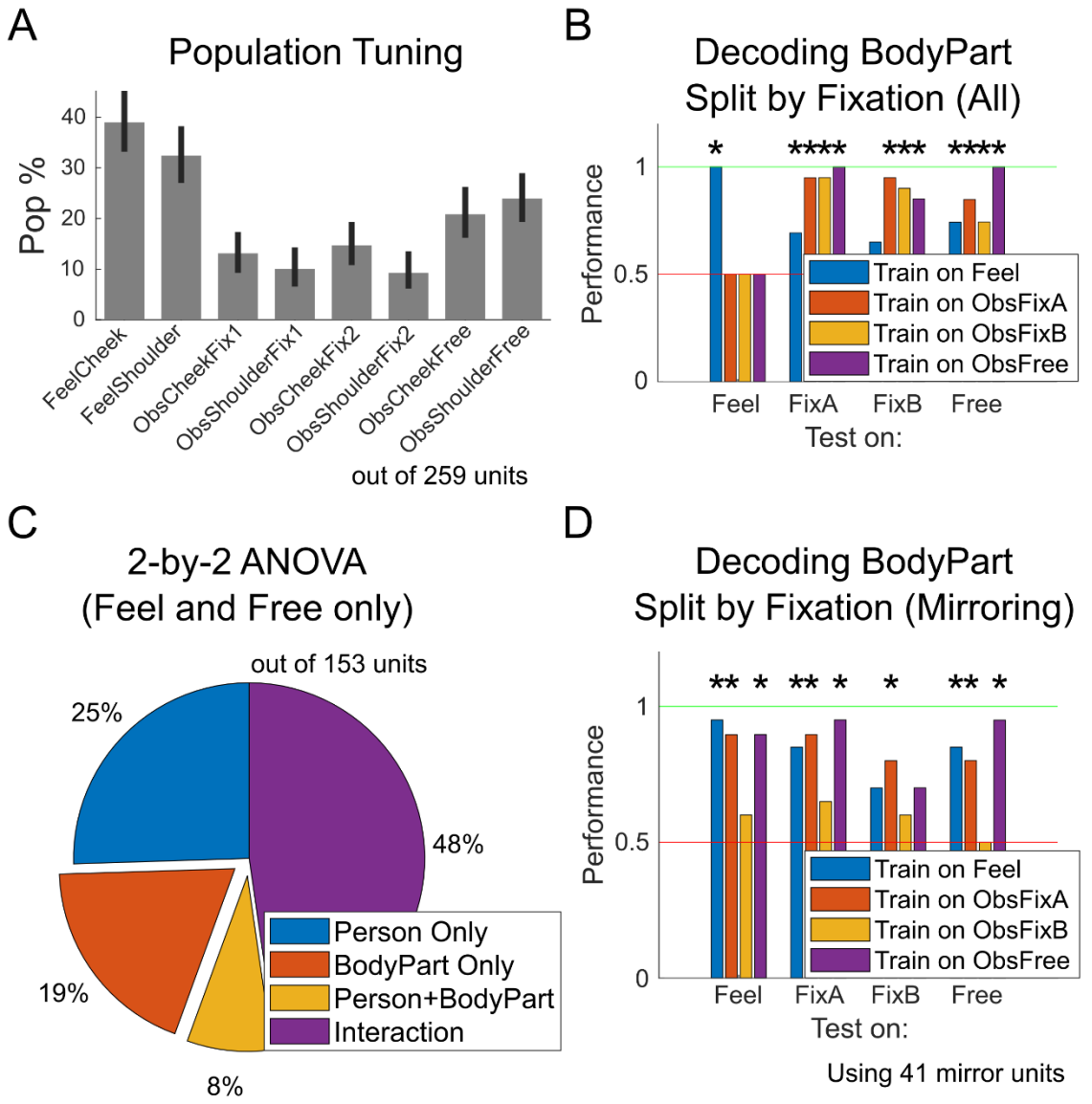

Figure 5.7. Effect of Fixation Condition (Free Gaze vs Fixation)

(A) Percent of the population significantly tuned to each of the feel conditions and fixation/free gaze conditions (bootstrap 95% CI, $p < 0.05$, FDR corrected). (B) Cross-decode analysis performed on the entire neural population, similar to the one in Figure 5.3C, testing generalization of body part representations across felt sensations and fixation/free gaze. (C) Pie chart similar to the one in Figure 5.2B, showing the unit tuning types as determined by a 2-by-2 ANOVA. Only felt and free gaze conditions were included in the analysis and only units with tuning to at least one of the terms in the ANOVA were included in the pie chart. (D) Cross-decode analysis, similar to (B), but performed on only the body part specific mirror units (i.e., the ones in the yellow/red sections in (C)). (Feel = felt, Obs = observed; FixA = fixation point above action, FixB = fixation point below action, Free = free gaze).

5.4 Discussion

5.4.1 Mixed Coding and Sensory Mirroring

In this study, we tested for sensory mirroring responses in human putative AIP. Specifically, we compared neural responses to felt and observed sensations on the cheek and shoulder, and found a strong body part specific mirroring component at both the single unit and population level.

A significant proportion of the population was tuned to each of the four conditions, with partially anatomically overlapping networks (Figure 5.1C). This is suggestive of a mixed coding structure and is consistent with our previous study that found mixed selectivity of variables in AIP (Zhang et al., 2017). Thus, the results of this current study further extend our findings on mixed selectivity in AIP, finding that besides the motor-related variables tested in our previous work (cognitive strategy, body part, and body side), the person variable, distinguishing between felt and observed sensations are also encoded in a mixed fashion.

Also consistent with our previous study, we found that the person variable was functionally segregated by the body part variable (Figure 5.3). This finding provides additional evidence of partially mixed selectivity as a mechanism for how AIP encodes different variables. This result also suggests a potential implementation mechanism for the sensory mirror system. Sensory mirroring is defined as observed sensations being processed and represented similarly to experienced sensations of the same body part. Functional segregation of person by body part, on the other hand, means that representations of felt/observed sensations are more similar within the same body part than across different body parts. The greater similarity of

felt/observed sensation representations across matched body parts is the very definition of sensory mirroring. In other words, the functional segregation of person by body part directly hints at the presence of a sensory mirroring response. Note, however, that we could not assume beforehand that the person variable would be functionally segregated by body part. The functional segregation of some variables by body part (as found in Zhang et al. 2017) does not mean that all variables would also be similarly segregated.

Furthermore, the presence of a mirror response does not require functional segregation by body parts. The similarity between representations of felt and observed sensations matched by body part could be high, independent of the similarity between the body part representations. Mirror responses could exist in other encoding schemes, as well. For example, if there were separate, non-overlapping populations tuned to each body part but invariant for felt vs observed sensations, we would still see a mirror response. Mirror neurons could also just exist independently in a randomly mixed encoding scheme. In that case, units tuned similarly to felt and observed sensations of the same body part would exist even by chance, without any need for functional segregation by body parts.

Different from the results in Zhang et al. 2017, however, we found that the correlation between representations differing only by body part is significantly greater than zero (Figure 5.3B and Figure 5.5CD), rather than orthogonal (Figure 5.6A in Zhang et al. 2017). While representations of person did not generalize perfectly across different body parts, they did still somewhat generalize (Figure 5.3D), suggesting imperfect functional segregation. This is actually consistent with the theory (also presented in our previous study) that the degree of correlation between

different representations is related to the amount of shared computations or the amount of learning transfer reasonably desirable between the conditions. In that study, we looked at imagined/attempted movements of the left/right hand/shoulder (hand squeezes and shoulder shrugs). In that context, the shared computations in the movements would be relatively minimal and it would be undesirable for any learning adaptations made from one body part's movements to transfer to the other, and thus the representations were uncorrelated with each other. In the context of this current study where we are comparing felt and observed sensations on the cheek and shoulder, however, it would be different. It would make sense for there to be some shared computations in the observation/feeling process and for some parts of the sensation representations (e.g., the rate of the touches, the force behind them, etc.) to generalize across different body parts (i.e., have a significant non-zero correlation).

In this study, we primarily tested sensations performed on two body parts, the cheek and the shoulder. These two body parts were chosen because they were fairly differentiable in representation, based on a tactile receptive field test (Figure 5.4B). However, the significant fraction of units showing only a main effect of person (Figure 5.2B, blue section) indicate a population of units that represent different body parts similarly. The straightforward interpretation is that these units are person-specific, coding person independently of body part. However, it is also possible that these units only appear invariant because of the specific body parts tested. The inclusion of other body parts in the analysis might reveal some of them to actually have some body part specificity. The analysis on the number of units invariant for a comparison of N body part receptive fields suggests that this is indeed possible (Figure 5.4C). In that analysis, we found that as we introduced more body parts, the number of units tuned and invariant to all of the body parts decreased drastically. Although that analysis was only

performed for felt sensations, the main point is that units invariant to two body parts may actually be differentially tuned to a third body part instead. In other words, many of the units found to be “person-specific” are not truly person-specific.

In terms of the mixing of other variables, we also compared the representations of different actions performed on the cheek and shoulder (Figure 5.5). Our recorded population was still tuned to the different observed actions (Figure 5.5AB, observe conditions), consistent with other studies on action observation in AIP (Nelissen et al., 2011; Pani et al., 2014). However, the differences in representation caused by different actions were minor relative to the larger representational differences caused by the actions being performed on different people or body parts (Figure 5.5CD). This suggests that, compared to the specific action being performed, our neural population encodes who or what is being touched more strongly. This is also consistent with past studies on AIP being selective for visual properties of an object to be grasped (Klaes et al., 2015; Murata et al., 2000).

5.4.2 Differences between Live Action and Video

For the main task used in this study, we presented all observed sensations in live action rather than in video, and with free gaze rather than with required eye fixation. Although we found a significant sensory mirroring response when using both presentation formats, we interestingly found more units tuned to the live action format than the video format (Figure 5.6A). This is despite the fact that we matched the video stimuli to be from the same perspective and distance as the live action stimuli.

This observation is consistent with a study in premotor cortex (area F5) of NHPs that found that naturalistic (i.e., live action) stimuli elicited stronger neural responses in terms of both percent tuned and larger firing rates (Caggiano et al., 2011). Furthermore, they found that the responses to live action and video were highly correlated and specific to action stimuli (as opposed to non-action-related stimuli), suggesting that the neural responses to actions performed live are a stronger version of the responses to actions performed in video. Considering that F5 is a region receiving strong projections from AIP and the two have related functions (Luppino et al., 1999), it is not unreasonable for the tuning differences between formats found in F5 to also be found upstream in AIP.

As for why AIP might prefer the live action format to video, there are several possibilities. First, there have been several studies in the IPS and F5 of NHPs finding that mirror neurons can be selective for actions performed near the observer (i.e., in the peripersonal space) vs actions performed far away (i.e., in the extrapersonal space) (Caggiano et al., 2009; Ishida et al., 2010). While these studies generally defined peripersonal and extrapersonal by distance from the observer, they also found similar effects when a transparent barrier was placed between the observer and the actor. In the context of some actions being performed on video, it is reasonable for the video actions to be “extrapersonal”, i.e., obstructed by a “barrier” (the LCD screen).

When we tested generalization of body part representations across felt sensations, observed live sensations, and observed video sensations, we found significant generalization when focusing on only the mirroring units (Figure 5.6D). When we looked at the entire population,

however, body part representations generalized from felt sensations to observed live/sensations, but not in the reverse-direction (Figure 5.6B). This is in stark contrast to the bidirectional generalization we found across the entire population in the main sensory mirroring task (Figure 5.3C), and might call into question whether the responses are truly mirroring. For the purposes of sensory mirroring, however, the essential element is the generalization of body parts from felt sensations to observed sensations, i.e., that the way the neural population processes observed sensations is similar to the way it processes felt sensations. This is the crucial element in the interpretation of mirror responses as understanding another's sensations in terms of the self. The representations of observed sensations are also relatively less well controlled as they are potentially subject to other visual factors (e.g., trial to trial differences due to the action being performed live/with free gaze), making it plausible for the generalizability of observed sensations to felt sensations to change. Thus, despite the change, the generalization of felt sensations to observed sensations still demonstrates sensory mirroring in our neural population.

Note that in the task testing the effects of fixation, all of the observe conditions were also presented in video. Thus, the above considerations about the video format vs live action apply to the fixation effect results as well (Figure 5.7).

Altogether, our tests on the effects of presentation format and fixation requirements show that, although the neural population is affected by these variables, the overall generalization of body part representations from felt sensations to observed sensations is not (Figure 6BD and Figure

7BD, blue bars). In other words, the sensory mirroring responses we examine in this study are present irrespective of the differences in visual presentation of the sensations.

6 Conclusion

In this dissertation, we studied mixed selectivity in the anterior intraparietal area (AIP) of human posterior parietal cortex (PPC). We found a variety of variables represented within the neural population recorded by our array. The subpopulations representing the variables were highly anatomically overlapping, suggestive of "mixed selectivity", a coding mechanism where individual units are tuned to idiosyncratic combinations of variables. Mixed selectivity enables a relatively small population of neurons to represent many variables in a high dimensional way and is a hallmark of high-level associative areas (Fusi et al., 2016; Raposo et al., 2014; Rigotti et al., 2013; Zipser and Andersen, 1988). Instead of the random mixed selectivity past studies have described, however, we found that some variables were randomly mixed while others were not. Representations of body side and cognitive strategy were organized by body part (i.e., "functionally segregated"), in a structure we termed "partially mixed selectivity" (Chapter 3). Furthermore, we found that the structure of the representations was largely preserved between training and online control (as opposed to the representations collapsing into an effector-independent intention signal) and that the tested movement conditions could all be used for closed-loop cortical control in a brain-machine interface (Chapter 4). Finally, we studied sensory mirror neuron responses in human AIP, finding a population mirroring response. This population response could be explained within the framework of partially mixed selectivity, with the person variable (distinguishing between felt and observed sensations) functionally segregated by body part. This was analogous to how the body side and cognitive strategy variables were functionally segregated by body part in Chapter 3. These results further point to partially mixed selectivity as a way to encode many variables in a relatively small neural population.

The fact that we found such a large variety of variables coded within the small 4 x 4 mm patch of cortex covered by our recording array is exciting for future BMI technologies. We were able to decode the many variables and use them for BMI control. The ability to decode movements of a large extent of the body (e.g., the entire extent of the arm, on both the left and right sides) as well as other potentially relevant variables (e.g., imagine vs attempt) is highly desirable, as it would reduce the number of implants/arrays necessary in a full-body BMI.

This ability is enabled by mixed selectivity, a property of high-level association brain areas such as PPC. Besides being a high-level area and mixed selective, the advantage of PPC is that it is a brain region involved in movement planning, intentions, goals, and decision-making (Aflalo et al., 2015; Andersen and Buneo, 2002; Andersen and Cui, 2009; Christopoulos et al., 2015; Mulliken et al., 2008a; Quiroga et al., 2006). These are all variables useful in developing a high-level BMI. Primary motor cortex (M1) is another brain area often used for BMI control applications, due to its representation of relative motor variables such as velocity (Georgopoulos et al., 1982; Hochberg et al., 2012; Sergio et al., 2005). Compared to PPC, however, M1 is a much more anatomically specialized region and a relatively lower-level brain area closer to the output stages of the motor pathway (Holdefer and Miller, 2002; Morrow et al., 2007; Morrow and Miller, 2003; Rathelot and Strick, 2009). As a result, we would expect the networks representing different body parts to be much less overlapping in M1 compared to PPC (if overlapping at all). The limited overlap means fewer neurons are tuned to multiple variables simultaneously (i.e., less mixing of variables), which in turn, might make it difficult to decode the variety of variables we were able to decode here in AIP (see also Mixed Coding in Chapter 2).

Although we have focused on AIP in this dissertation, we would expect other high-level brain areas to exhibit partially mixed selectivity, too, with some variables randomly mixed and others functionally segregated. In Chapter 5, for example, we found observed and felt sensations to be functionally segregated by body part, with the body part variable being specially coded in our population. This makes sense in the context of AIP being related to motor intentions, and thus potentially needing to clearly distinguish between body parts. In contrast, consider inferotemporal cortex, a region that has been implicated in object and facial identity recognition and thought to be more concerned with identity or self vs other (Chang and Tsao, 2017; Hung et al., 2005; Sliwa et al., 2016; Tsao et al., 2006; Zhang et al., 2011). As a result, the brain region might have representations functionally segregated by person, coding who was touched more specially and distinguishing more between different objects and identities instead.

The specialization of a brain area is traditionally defined by what variables it encodes. The encoding of many different variables in higher-level associative brain areas, however, complicates this definition. Our discovery of partially mixed selectivity in AIP suggests an alternative definition, where the specialization of associative brain areas can be defined by which variables are functionally segregating/most strongly represented. In the case of AIP, for example, we might think of it as caring more about body part representations, even though many other variables are also represented.

The structure of partially mixed selectivity also has some potential applications from an engineering perspective. In AIP, for example, knowing that AIP has body parts functionally segregated, we could potentially regularize BMI decoders to comport with or take advantage of

the known structure (Lu et al., 2000; Maruyama and Shikita, 2014; Oktay et al., 2017). For example, instead of training a neural network by searching the space of all possible combinations of connections/weights, we can constrain the space to focus on those having partially overlapping networks and having weights exhibiting correlation patterns consistent with those found in Figure 3.5 and Figure 3.6. The partially mixed structure could also be used to initially seed a decoder and speed up the training process. For example, a decoder using imagined left hand movements could be initialized with weights based on those from an already-trained decoder that uses imagined right hand movements, since we know movement representations can generalize from one hand to the other (Figure 3.7F). Using these initial weights would start the decoder training process closer to an optimal solution than simply using random weights (Thrun and O'Sullivan, 1998; Yosinski et al., 2014). Decoders could also exploit the partially mixed selectivity structure using more structured methods (Bansal et al., 2004; Li et al., 2014; Salakhutdinov et al., 2013; Silla and Freitas, 2011; Zhou et al., 2011). For example, instead of trying to differentiate the motor intentions of the left/right hand/shoulder directly, a decoder might take a more hierarchical approach and first decode the body part (i.e., the organizing variable) before decoding specific aspects of the identified body part. These examples could extend to other brain areas and variables as well, depending on the specifics of the partially mixed structure.

Altogether, the work presented in this dissertation has broad applications in furthering brain-machine interfaces, decoding algorithms, and our understanding of the structure and mechanisms of the brain.

7 References

Adibi, M., McDonald, J.S., Clifford, C.W.G., and Arabzadeh, E. (2014). Population Decoding in Rat Barrel Cortex: Optimizing the Linear Readout of Correlated Population Responses. PLoS Computational Biology *10*, e1003415.
Adolphs, R. (2009). The Social Brain: Neural Basis of Social Knowledge. Annual Review of Psychology *60*, 693-716.
Adrian, E.D., and Zotterman, Y. (1926). The impulses produced by sensory nerve-endings. The Journal of Physiology *61*, 151-171.
Aflalo, T., Kellis, S., Klaes, C., Lee, B., Shi, Y., Pejsa, K., Shanfield, K., Hayes-Jackson, S., Aisen, M., Heck, C.*, et al.* (2015). Decoding motor imagery from the posterior parietal cortex of a tetraplegic human. Science *348*, 906-910.
Ajiboye, A.B., Willett, F.R., Young, D.R., Memberg, W.D., Murphy, B.A., Miller, J.P., Walter, B.L., Sweet, J.A., Hoyen, H.A., Keith, M.W.*, et al.* (2017). Restoration of reaching and grasping movements through brain-controlled muscle stimulation in a person with tetraplegia: a proof-of-concept demonstration. The Lancet *389*, 1821-1830.
Amemiya, K., Ishizu, T., Ayabe, T., and Kojima, S. (2010). Effects of motor imagery on intermanual transfer: A near-infrared spectroscopy and behavioral study. Brain Research *1343*, 93-103.
Andersen, R.A., Asanuma, C., Essick, G., and Siegel, R.M. (1990). Corticocortical connections of anatomically and physiologically defined subdivisions within the inferior parietal lobule. J Comp Neurol *296*, 65-113.
Andersen, R.A., and Buneo, C.A. (2002). Intentional maps in posterior parietal cortex. Annual Review of Neuroscience *25*, 189-220.
Andersen, R.A., and Cui, H. (2009). Intention, Action Planning, and Decision Making in Parietal-Frontal Circuits. Neuron *63*, 568-583.
Andersen, R.A., Essick, G.K., and Siegel, R.M. (1987). Neurons of area 7 activated by both visual stimuli and oculomotor behavior. Experimental Brain Research *67*, 316-322.
Andersen, R.A., Essick, G.K., and Siegel, R.N. (1985). Encoding of spatial location by posterior parietal neurons. Science *230*, 456-458.
Andersen, R.A., Kellis, S., Klaes, C., and Aflalo, T. (2014). Toward more versatile and intuitive cortical brain-machine interfaces. Curr Biol *24*, R885-897.
Andersen, R.A., and Mountcastle, V.B. (1983). The influence of the angle of gaze upon the excitability of the light-sensitive neurons of the posterior parietal cortex. J Neurosci *3*, 532-548.
Anderson, W.S., and Kreiman, G. (2011). Neuroscience: What we cannot model, we do not understand. Curr Biol *21*, R123-125.
Asher, I., Stark, E., Abeles, M., and Prut, Y. (2007). Comparison of direction and object selectivity of local field potentials and single units in macaque posterior parietal cortex during prehension. Journal of neurophysiology *97*, 3684-3695.
Astafiev, S.V., Shulman, G.L., Stanley, C.M., Snyder, A.Z., Van Essen, D.C., and Corbetta, M. (2003). Functional organization of human intrapaietal and frontal cortex for attending, looking, and pointing. Journal of Neuroscience *23*, 4689-4699.
Avillac, M., Deneve, S., Olivier, E., Pouget, A., and Duhamel, J.R. (2005). Reference frames for representing visual and tactile locations in parietal cortex. Nat Neurosci *8*, 941-949.
Balasubramanian, K., Vaidya, M., Southerland, J., Badreldin, I., Eleryan, A., Takahashi, K., Qian, K., Slutzky, M.W., Fagg, A.H., Oweiss, K.*, et al.* (2017). Changes in cortical network connectivity

with long-term brain-machine interface exposure after chronic amputation. Nat Commun *8*, 1796.
Baldauf, D., Cui, H., and Andersen, R.A. (2008). The Posterior Parietal Cortex Encodes in Parallel Both Goals for Double-Reach Sequences. Journal of Neuroscience *28*, 10081-10089.
Balint, R. (1909). Seelenlahmung des "Schauens," optische Ataxie, raumliche Storung der Aufmerksamkeit. Monatsschr Psychiatr Neurol *25*, 51-81.
Banissy, M., and Ward, J. (2013). Mechanisms of self-other representations and vicarious experiences of touch in mirror-touch synesthesia. Frontiers in Human Neuroscience *7*.
Banissy, M.J., Cohen Kadosh, R., Maus, G.W., Walsh, V., and Ward, J. (2009). Prevalence, characteristics and a neurocognitive model of mirror-touch synaesthesia. Exp Brain Res *198*, 261-272.
Banissy, M.J., and Ward, J. (2007). Mirror-touch synesthesia is linked with empathy. Nat Neurosci *10*, 815-816.
Bansal, N., Blum, A., and Chawla, S. (2004). Correlation Clustering. Machine Learning *56*, 89-113.
Batista, A.P., Buneo, C.A., Snyder, L.H., and Andersen, R.A. (1999). Reach plans in eye-centered coordinates. Science *285*, 257-260.
Bellman, R.E. (2015). Adaptive control processes: a guided tour (Princeton university press).
Benedek, M., Jauk, E., Beaty, R.E., Fink, A., Koschutnig, K., and Neubauer, A.C. (2016). Brain mechanisms associated with internally directed attention and self-generated thought. Scientific Reports *6*, 22959.
Beurze, S.M., de Lange, F.P., Toni, I., and Medendorp, W.P. (2009). Spatial and effector processing in the human parietofrontal network for reachces and saccades. J Neurophsiol *101*, 3053-3062.
Bhagat, N.A., Venkatakrishnan, A., Abibullaev, B., Artz, E.J., Yozbatiran, N., Blank, A.A., French, J., Karmonik, C., Grossman, R.G., O'Malley, M.K*., et al.* (2016). Design and Optimization of an EEG-Based Brain Machine Interface (BMI) to an Upper-Limb Exoskeleton for Stroke Survivors. Frontiers in Neuroscience *10*, 122.
Bisley, J.W., and Goldberg, M.E. (2010). Attention, Intention, and Priority in the Parietal Lobe. Annual review of neuroscience *33*, 1-21.
Blakemore, S.J., Bristow, D., Bird, G., Frith, C., and Ward, J. (2005). Somatosensory activations during the observation of touch and a case of vision–touch synaesthesia. Brain *128*, 1571-1583.
Blokland, Y., Spyrou, L., Thijssen, D., Eijsvogels, T., Colier, W., Floor-Westerdijk, M., Vlek, R., Bruhn, J., and Farquhar, J. (2014). Combined EEG-fNIRS Decoding of Motor Attempt and Imagery for Brain Switch Control: An Offline Study in Patients With Tetraplegia. IEEE Transactions on Neural Systems and Rehabilitation Engineering *22*, 222-229.
Brainard, D.H. (1997). The psychophysics toolbox. Spatial Vision *10*, 433-436.
Brotchie, P.R., Andersen, R.A., Snyder, L.H., and Goodman, S.J. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. Nature *375*, 232-235.
Buonomano, D.V., and Maass, W. (2009). State-dependent computations: spatiotemporal processing in cortical networks. Nat Rev Neurosci *10*, 113-125.
Buzsáki, G., Anastassiou, C.A., and Koch, C. (2012). The origin of extracellular fields and currents - EEG, ECog, LFP, and spikes. Nat Rev Neurosci *13*, 407-420.
Caggiano, V., Fogassi, L., Rizzolatti, G., Pomper, J.K., Thier, P., Giese, M.A., and Casile, A. (2011). View-based encoding of actions in mirror neurons of area f5 in macaque premotor cortex. Curr Biol *21*, 144-148.
Caggiano, V., Fogassi, L., Rizzolatti, G., Thier, P., and Casile, A. (2009). Mirror Neurons Differentially Encode the Peripersonal and Extrapersonal Space of Monkeys. Science *324*, 403-406.

Caramazza, A., Anzellotti, S., Strnad, L., and Lingnau, A. (2014). Embodied Cognition and Mirror Neurons: A Critical Assessment. Annual Review of Neuroscience *37*, 1-15.
Chang, L., and Tsao, D.Y. (2017). The Code for Facial Identity in the Primate Brain. Cell *169*, 1013-1028.e1014.
Chang, S.W.C., Dickinson, A.R., and Snyder, L.H. (2008). Limb-Specific Representation for Reaching in the Posterior Parietal Cortex. The Journal of Neuroscience *28*, 6128-6140.
Chang, S.W.C., and Snyder, L.H. (2012). The representations of reach endpoints in posterior parietal cortex depend on which hand does the reaching. Journal of neurophysiology *107*, 2352-2365.
Chase, S.M., Schwartz, A.B., and Kass, R.E. (2009). Bias, optimal linear estimation, and the differences between open-loop simulation and closed-loop performance of spiking-based brain-computer interface algorithms. Neural networks : the official journal of the International Neural Network Society *22*, 1203-1213.
Chaudhary, U., Birbaumer, N., and Curado, M.R. (2015). Brain-Machine Interface (BMI) in paralysis. Annals of Physical and Rehabilitation Medicine *58*, 9-13.
Chestek, C.A., Gilja, V., Nuyujukian, P., Foster, J.D., Fan, J.M., Kaufman, M.T., Churchland, M.M., Rivera-Alvidrez, Z., Cunningham, J.P., Ryu, S.I.*, et al.* (2011). Long-term stability of neural prosthetic control signals from silicon cortical arrays in rhesus macaque motor cortex. Journal Of Neural Engineering *8*, 045005.
Cho, M.J., Oh, C.W., Kwon, O.K., Byoun, H.S., Lee, S.U., Kim, T., Chung, Y.S., Ban, S.P., and Bang, J.S. (2017). Comparison of Unilateral and Bilateral Craniotomy for the Treatment of Bilateral Middle Cerebral Artery Aneurysms: Anatomic and Clinical Parameters and Surgical Outcomes. World Neurosurgery *108*, 627-635.
Chong, T.T.J., Cunnington, R., Williams, M.A., Kanwisher, N., and Mattingley, J.B. (2008). fMRI Adaptation Reveals Mirror Neurons in Human Inferior Parietal Cortex. Current biology : CB *18*, 1576-1580.
Christopoulos, N.V., Bonaiuto, J., Kagan, I., and Andersen, R.A. (2015). Inactivation of parietal reach region affects reaching but not saccade choices in internally guided decisions. Journal of Neuroscience *35*, 11719-11728.
Churchland, M.M., and Cunningham, J.P. (2015). A Dynamical Basis Set for Generating Reaches. Cold Spring Harbor Symposia on Quantitative Biology, 024703.
Churchland, M.M., Cunningham, J.P., Kaufman, M.T., Ryu, S.I., and Shenoy, K.V. (2010). Cortical preparatory activity: representation of movement or first cog in a dynamical machine? Neuron *68*, 387-400.
Collinger, J.L., Wodlinger, B., Downey, J.E., Wang, W., Tyler-Kabara, E.C., Weber, D.J., McMorland, A.J.C., Velliste, M., Boninger, M.L., and Schwartz, A.B. (2013). High-performance neuroprosthetic control by an individual with tetraplegia. Lancet *381*, 557-564.
Connolly, J.D., Andersen, R.A., and Goodale, M.A. (2003). FMRI evidence for a 'parietal reach region' in the human brain. Exp Brain Res *153*, 140-145.
Cui, H., and Andersen, R.A. (2007). Posterior parietal cortex encodes autonomously selected motor plans. Neuron *56*, 552-559.
Culham, J.C., Danckert, S.L., DeSouza, J.F.X., Gati, J.S., Menon, R.S., and Goodale, M.A. (2003). Visaully guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. Exp Brain Res *153*, 180-189.
Cunningham, J.P., Nuyujukian, P., Gilja, V., Chestek, C.A., Ryu, S.I., and Shenoy, K.V. (2011). A closed-loop human simulator for investigating the role of feedback control in brain-machine interfaces. Journal of neurophysiology *105*, 1932-1949.

Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., and Rizzolatti, G. (1992). Understanding Motor Events - a Neurophysiological Study. Experimental Brain Research *91*, 176-180.
Dickstein, R. (2007). Motor Imagery in Physical Therapy Practice. Journal of the American Physical Therapy Association *87*, 942-953.
Fattori, P., Breveglieri, R., Marzocchi, N., Filippini, D., Bosco, A., and Galletti, C. (2009). Hand orientation during reach-to-grasp movements modulates neuronal activity in the medial posterior parietal area V6A. The Journal of Neuroscience *29*, 1928-1936.
Filimon, F., Nelson, J.D., Hagler, D.J., and Sereno, M.I. (2007). Human cortical representations for reaching: Mirror neurons for execution, observation, and imagery. NeuroImage *37*, 1315-1328.
Filimon, F., Rieth, C.A., Sereno, M.I., and Cottrell, G.W. (2015). Observed, Executed, and Imagined Action Representations can be Decoded From Ventral and Dorsal Areas. Cerebral Cortex *25*, 3144-3158.
Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., and Rizzolatti, G. (2005). Parietal Lobe: From Action Organization to Intention Understanding. Science *308*, 662-667.
Foundation for Spinal Cord Injury Prevention, C.C. (2009). Spinal Cord Injury Facts. In Spinal Cord Injury Facts.
Fritsch, G., and Hitzig, E. (1960). Ueber die elektrishe Erregarkeit des Grosshirns (On the electrical excitability of the cerebrum). Some Papers on the Cerebral Cortex (Translated by G von Bonin), 73-96.
Fujii, N., Hihara, S., and Iriki, A. (2008). Social cognition in premotor and parietal cortex. Social Neuroscience *3*, 250-260.
Fusi, S., Miller, E.K., and Rigotti, M. (2016). Why neurons mix: high dimensionality for higher cognition. Current Opinion In Neurobiology *37*, 66-74.
Gail, A., and Andersen, R.A. (2006). Neural dynamics in monkey parietal reach region reflect context-specific sensorimotor transformations. J Neurosci *26*, 9376-9384.
Gallese, V., Fadiga, L., Fogassi, L., and Rizzolatti, G. (1996). Action recognition in the premotor cortex. Brain *119*, 593-609.
Gallese, V., Murata, A., Kaseda, M., Niki, N., and Sakata, H. (1994). Deficit of hand preshaping after muscimol injection in monkey parietal cortex. Neuroreport *5*, 1525-1529.
Gallivan, J.P., McLean, D.A., Smith, F.W., and Culham, J.C. (2011). Decoding effector-dependent and effector-independent movement intentions from human parieto-frontal brain activity. The Journal of Neuroscience *31*, 17149-17168.
Gallivan, J.P., McLean, D.A., Flanagan, J.R., and Culham, J.C. (2013). Where one hand meets the other: limb-specific and action-dependent movement plans decoded from preparatory signals in single human frontoparietal brain areas. Journal of Neuroscience *33*, 1991-2008.
Ganguly, K., and Carmena, J.M. (2010). Neural correlates of skill acquisition with a cortical brain-machine interface. Journal of motor behavior *42*, 355-360.
Gazzola, V., Aziz-Zadeh, L., and Keysers, C. (2006). Empathy and the Somatotopic Auditory Mirror System in Humans. Current Biology *16*, 1824-1829.
Gentile, G., Petkova, V.I., and Ehrsson, H.H. (2011). Integration of Visual and Tactile Signals From the Hand in the Human Brain: An fMRI Study. Journal of neurophysiology *105*, 910-922.
Georgopoulos, A.P., Kalaska, J.F., Caminiti, R., and Massey, J.T. (1982). On the relations between the direction of two-dimensional arm movements into cell discharge in primate motor cortex. Journal of Neuroscience *2*, 1527-1537.
Gerardin, E., Sirigu, A., Lehéricy, S., Poline, J.B., Gaymard, B., Marsault, C., Agid, Y., and Le Bihan, D. (2000). Partially overlapping neural networks for real and imagined hand movements. Cerebral Cortex *10*, 1093-1104.

Glover, G.H. (2011). Overview of Functional Magnetic Resonance Imaging. Neurosurgery clinics of North America *22*, 133-139.
Graf, B.A.A., and Andersen, R.A. (2014). Brain-machine interface for eye movements. PNAS *111*, 17630-17635.
Graziano, M.S.A., and Aflalo, T.N. (2007). Rethinking cortical organization: moving away from discrete areas arranged in hierarchies. The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry *13*, 138-147.
Grefkes, C., and Fink, G.R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. Journal of Anatomy *207*, 3-17.
Grèzes, J., Armony, J.L., Rowe, J., and Passingham, R.E. (2003). Activations related to "mirror" and "canonical" neurones in the human brain: an fMRI study. NeuroImage *18*, 928-937.
Harris, K.D., Henze, D.A., Csicsvari, J., Hirase, H., and Buzsaki, G. (2000). Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. J Neurophysiol *84*, 401-414.
Harris, K.D., Quiroga, R.Q., Freeman, J., and Smith, S.L. (2016). Improving data quality in neuronal population recordings. Nat Neurosci *19*, 1165-1174.
Heed, T., Beurze, S.M., Toni, I., Roder, B., and Medendorp, W.P. (2011a). Functional rather than effector-specific organization of human posterior parietal cortex. Journal of Neuroscience *31*, 3066-3076.
Heyes, C. (2010). Where do mirror neurons come from? Neurosci Biobehav Rev *34*, 575-583.
Hickok, G. (2009). Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans. Journal of cognitive neuroscience *21*, 1229-1243.
Hinkley, L.B.N., Krubitzer, L.A., Padberg, J., and Disbrow, E.A. (2009). Visaul-manual exploration and posterior parietal cortex in humans. J Neurophsiol *102*, 3433-3446.
Hochberg, L.R., Bacher, D., Jarosiewicz, B., Masse, N.Y., Simeral, J.D., Vogel, J., Haddadin, S., Liu, J., Cash, S.S., Smagt, P.*, et al.* (2012). Reach and grasp by people with tetraplegia using a neurally controlled robotic arm. Nature *485*, 372-375.
Hochberg, L.R., Serruya, M.D., Friehs, G.M., Mukand, J.A., Saleh, M., Caplan, A.H., Branner, A., Chen, D., Penn, R.D., and Donoghue, J.P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature *442*, 164-171.
Höhne, J., Holz, E., Staiger-Sälzer, P., Müller, K.-R., Kübler, A., and Tangermann, M. (2014). Motor Imagery for Severely Motor-Impaired Patients: Evidence for Brain-Computer Interfacing as Superior Control Solution. PloS one *9*, e104854.
Holdefer, R.N., and Miller, L.E. (2002). Primary motor cortical neurons encode functional muscle synergies. Experimental Brain Research *146*, 233-243.
Holmes, G. (1918). Disturbances of visual orientation. Br J Opthalmol *2*, 449-468.
Hung, C.P., Kreiman, G., Poggio, T., and DiCarlo, J.J. (2005). Fast Readout of Object Identity from Macaque Inferior Temporal Cortex. Science *310*, 863-866.
Hwang, E., Hauschild, M., Wilke, M., and Andersen, R. (2012). Inactivation of the parietal reach region causes optic ataxia, impairing reaches but not saccades. Neuron *76*, 1021-1029.
Hwang, E., Hauschild, M., Wilke, M., and Andersen, R. (2014). Spatial and Temporal Eye-Hand Coordination Relies on the Parietal Reach Region. Journal of Neuroscience *34*, 12884-12892.
Iacoboni, M. (2009). Imitation, empathy, and mirror neurons. Annu Rev Psychol *60*, 653-670.
Ifft, P.J., Shokur, S., Li, Z., Lebedev, M.A., and Nicolelis, M.A.L. (2013). Brain-Machine Interface Enables Bimanual Arm Movements in Monkeys. Sci Transl Med *5*.
Ishida, H., Nakajima, K., Inase, M., and Murata, A. (2010). Shared Mapping of Own and Others' Bodies in Visuotactile Bimodal Area of Monkey Parietal Cortex. Journal of Cognitive Neuroscience *22*, 83-96.

Jaeger, H., and Haas, H. (2004). Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. Science *304*, 78-80.
Jarosiewicz, B., Sarma, A.A., Bacher, D., Masse, N.Y., Simeral, J.D., Sorice, B., Oakley, E.M., Blabe, C., Pandarinath, C., Gilja, V*., et al.* (2015). Virtual typing by people with tetraplegia using a self-calibrating intracortical brain-computer interface. Sci Transl Med *7*.
Jastorff, J., Begliomini, C., Fabbri-Destro, M., Rizzolatti, G., and Orban, G.A. (2010). Coding Observed Motor Acts: Different Organizational Principles in the Parietal and Premotor Cortex of Humans. J Neurophysiol *104*, 128-140.
Jeannerod, M. (1995). Mental imagery in the motor context. Neuropsychologia *33*, 1419-1432.
Kaufman, M.T., Churchland, M.M., Ryu, S.I., and Shenoy, K.V. (2014). Cortical activity in the null space: permitting preparation without movement. Nat Neurosci *17*, 440-448.
Keysers, C., and Gazzola, V. (2006). Towards a unifying neural theory of social cognition. In Progress in Brain Research, S. Anders, G. Ende, M. Junghofer, J. Kissler, and D. Wildgruber, eds. (Elsevier), pp. 379-401.
Keysers, C., Kohler, E., Umiltà, M.A., Nanetti, L., Fogassi, L., and Gallese, V. (2003). Audiovisual mirror neurons and action recognition. Experimental Brain Research *153*, 628-636.
Keysers, C., Wicker, B., Gazzola, V., Anton, J.-L., Fogassi, L., and Gallese, V. (2004). A Touching Sight: SII/PV Activation during the Observation and Experience of Touch. Neuron *42*, 335-346.
Kilner, J.M., Neal, A., Weiskopf, N., Friston, K.J., and Frith, C.D. (2009). Evidence of Mirror Neurons in Human Inferior Frontal Gyrus. The Journal of Neuroscience *29*, 10153-10159.
Kim, S.-P., Simeral, J.D., Hochberg, L.R., Donoghue, J.P., and Black, M.J. (2008). Neural control of computer cursor velocity by decoding motor cortical spiking activity in humans with tetraplegia. Journal Of Neural Engineering *5*, 455-476.
Klaes, C., Kellis, S., Aflalo, T., Lee, B., Pejsa, K., Shanfield, K., Hayes-Jackson, S., Aisen, M., Heck, C., Liu, C*., et al.* (2015). Hand Shape Representations in the Human Posterior Parietal Cortex. Journal of Neuroscience *35*, 15466-15476.
Klaes, C., Kellis, S., Aflalo, T., Lee, B., Pejsa, K. Shanfield, K., Hayes-Jackson, S., Aisen, M., Heck, C., Liu, C. and Andersen, R.A. (2015). Hand shape representations in the human posterior parietal cortex. J Neurosci *35*, 15466-15476.
Kubanek, J., and Snyder, L.H. (2015). Reward-base decision signals in parietal cortex are partially embodied. Journal of Neuroscience *35*, 4869-4881.
Kuhtz-Buschbeck, J.P., Sundholm, L.K., Eliasson, A.-C., and Forssberg, H. (2000). Quantitative assessment of mirror movements in children and adolescents with hemiplegic cerebral palsy. Developmental Medicine & Child Neurology *42*, 728-736.
Lehmann, S.J., and Scherberger, H. (2013). Reach and Gaze Representations in Macaque Parietal and Premotor Grasp Areas. The Journal of neuroscience : the official journal of the Society for Neuroscience *33*, 7038-7049.
Lehmann, S.J., and Scherberger, H. (2015). Spatial Representations in Local Field Potential Activity of Primate Anterior Intraparietal Cortex (AIP). PloS one *10*, e0142679.
Leuthardt, E.C., Schalk, G., Wolpaw, J.R., Ojemann, J.G., and Moran, D.W. (2004). A brain-computer interface using electrocorticographic signals in humans. J Neural Eng *1*, 63-71.
Levy, I., Schluppeck, D., Heeger, D.J., and Glimcher, P.W. (2007). Specificity of human cortical areas for reaches and saccades. Journal of Neuroscience *27*, 4687-4696.
Li, C.S., and Andersen, R.A. (2001). Inactivation of macaque lateral intraparietal area delays initiation of the second saccade predominantly from contralesional eye positions in a double-saccade task. Exp Brain Res *137*, 45-57.

Li, Z., Liu, J., Yang, Y., Zhou, X., and Lu, H. (2014). Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection. IEEE Transactions on Knowledge & Data Engineering *26*, 2138-2150.
Lindsay, G.W., Rigotti, M., Warden, M.R., Miller, E.K., and Fusi, S. (2017). Hebbian Learning in a Random Network Captures Selectivity Properties of the Prefrontal Cortex. The Journal of Neuroscience *37*, 11021-11036.
Liu, J., Harris, A., and Kanwisher, N. (2010). Perception of Face Parts and Face Configurations: An fMRI Study. Journal of cognitive neuroscience *22*, 203-211.
López-Larraz, E., Antelis, J.M., Montesano, L., Gil-Agudo, A., and Minguez, J. (2012). Continuous decoding of motor attempt and motor imagery from EEG activity in spinal cord injury patients. Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society IEEE Engineering in Medicine and Biology Society Conference *2012*, 1798-1801.
Lotze, M., Erb, M., Flor, H., Huelsmann, E., Godde, B., and Grodd, W. (2000). fMRI Evaluation of Somatotopic Representation in Human Primary Motor Cortex. NeuroImage *11*, 473-481.
Lu, B., Hirasawa, K., and Murata, J. (2000). A new learning method using prior information of neural networks. Artificial Life and Robotics *4*, 78-83.
Luppino, G., Murata, A., Govoni, P., and Matelli, M. (1999). Largely segregated parietofrontal connections linking rostral intraparietal cortex (areas AIP and VIP) and the ventral premotor cortex (areas F5 and F4). Exp Brain Res *128*, 181-187.
Maeda, K., Ishida, H., Nakajima, K., Inase, M., and Murata, A. (2015). Functional Properties of Parietal Hand Manipulation–related Neurons and Mirror Neurons Responding to Vision of Own Hand Action. Journal of Cognitive Neuroscience *27*, 560-572.
Majaj, N.J., Hong, H., Solomon, E.A., and DiCarlo, J.J. (2015). Simple Learned Weighted Sums of Inferior Temporal Neuronal Firing Rates Accurately Predict Human Core Object Recognition Performance. The Journal of Neuroscience *35*, 13402-13418.
Maruyama, O., and Shikita, S. (2014). A scale-free structure prior for Bayesian inference of Gaussian graphical models. Paper presented at: 2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).
Matsuzaka, Y., Picard, N., and Strick, P.L. (2007). Skill representation in the primary motor cortex after long-term practice. Journal of neurophysiology *97*, 1819-1832.
McKenzie, S., Frank, A.J., Kinsky, N.R., Porter, B.R., and Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. Neuron *83*, 202-215.
Meng, J., Zhang, S., Bekyo, A., Olsoe, J., Baxter, B., and He, B. (2016). Noninvasive Electroencephalogram Based Control of a Robotic Arm for Reach and Grasp Tasks. Scientific Reports *6*, 38565.
Morris, R.G. (1999). D.O. Hebb: The Organization of Behavior, Wiley: New York; 1949. Brain Res Bull *50*, 437.
Morrow, M.M., Jordan, L.R., and Miller, L.E. (2007). Direct comparison of the task-dependent discharge of M1 in hand space and muscle space. Journal of neurophysiology *97*, 1786-1798.
Morrow, M.M., and Miller, L.E. (2003). Prediction of Muscle Activity by Populations of Sequentially Recorded Primary Motor Cortex Neurons. Journal of neurophysiology *89*, 2279-2288.
Mountcastle, V.B., Lynch, J. C., Georgopoulos, A., Sakata, H., Acuna, C. (1975). Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. J Neurophysiol *38*, 871-908.

Mukamel, R., Ekstrom, A.D., Kaplan, J., Iacoboni, M., and Fried, I. (2010). Single neuron responses in humans during execution and observation of actions. Current biology : CB *20*, 750-756.
Mulliken, G.H., Musallam, S., and Andersen, R.A. (2008a). Decoding Trajectories from Posterior Parietal Cortex Ensembles. The Journal of Neuroscience *28*, 12913-12926.
Mulliken, G.H., Musallam, S., and Andersen, R.A. (2008b). Forward estimation of movement state in posterior parietal cortex. Proceedings of the National Academy of Sciences *105*, 8170-8177.
Munzert, J., Lorey, B., and Zentgraf, K. (2009). Cognitive motor processes: the role of motor imagery in the study of motor representations. Brain research reviews *60*, 306-326.
Murata, A., Gallese, V., Luppino, G., Kaseda, M., and Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. Journal of neurophysiology *83*, 2580-2601.
Murata, A., Gallese, V., Luppino, G., Kaseda, M., and Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. J Neurophysiol *83*, 2580-2601.
Musallam, S., Corneil, B., Greger, B., Scherberger, H., and Andersen, R.A. (2004). Cognitive control signals for neural prosthetics. Science *305*, 258-262.
Nelissen, K., Borra, E., Gerbella, M., Rozzi, S., Luppino, G., Vanduffel, W., Rizzolatti, G., and Orban, G.A. (2011). Action observation circuits in the macaque monkey cortex. The Journal of neuroscience : the official journal of the Society for Neuroscience *31*, 3743-3756.
Noachtar, S., and Remi, J. (2009). The role of EEG in epilepsy: a critical review. Epilepsy & behavior : E&B *15*, 22-33.
Nuyujukian, P., Kao, J.C., Ryu, S.I., and Shenoy, K.V. (2017). A Nonhuman Primate Brain-Computer Typing Interface. Proceedings of the IEEE *105*, 66-72.
Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M.P., Bai, W., Caballero, J., Guerrero, R., Cook, S.A., de Marvao, A., Dawes, T*., et al.* (2017). Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation. CoRR *abs/1705.08302*.
Osborn, J., and Derbyshire, S.W. (2010). Pain sensation evoked by observing injury in others. Pain *148*, 268-274.
Pandarinath, C., Gilja, V., Blabe, C.H., Nuyujukian, P., Sarma, A.A., Sorice, B.L., Eskandar, E.N., Hochberg, L.R., Henderson, J.M., and Shenoy, K.V. (2015). Neural population dynamics in human motor cortex during movements in people with ALS. Elife *4*.
Pani, P., Theys, T., Romero, M.C., and Janssen, P. (2014). Grasping Execution and Grasping Observation Activity of Single Neurons in the Macaque Anterior Intraparietal Area. Journal of Cognitive Neuroscience *26*, 2342-2355.
Penfield, W., and Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain *60*, 389-443.
Pesaran, B., Nelson, M.J., and Andersen, R.A. (2006). Dorsal premotor neurons encode the relative position of the hand, eye, and goal during reach planning. Neuron *51*, 125-134.
Pesaran, B., Nelson, M.J., and Andersen, R.A. (2010). A Relative Position Code for Saccades in Dorsal Premotor Cortex. The Journal of Neuroscience *30*, 6527-6537.
Pilgramm, S., de Haas, B., Helm, F., Zentgraf, K., Stark, R., Munzert, J., and Krüger, B. (2016). Motor imagery of hand actions: Decoding the content of motor imagery from brain activity in frontal and parietal motor areas. Human Brain Mapping *37*, 81-93.
Pouget, A., and Sejnowski, T.J. (1997). Spatial transformations in the parietal cortex using basis functions. J Cogn Neurosci *9*, 222-237.

Pouzat, C., Mazor, O., and Laurent, G. (2002). Using noise signature to optimize spike-sorting and to assess neuronal classification quality. Journal of Neuroscience Methods *122*, 43-57.
Prado, J., Clavagnier, S., Otzenberger, H., Scheiber, C., Kennedy, H. and Pererin, M.T. (2005). Two cortical systems for reaching in central and peripheral vision. Neuron *48*, 849-858.
Quiroga, R., Snyder, L., Batista, A.P., Cui, H., and Andersen, R.A. (2006). Movement intention is better predicted than attention in the posterior parietal cortex. The Journal of Neuroscience *26*, 3615-3620.
Raposo, D., Kaufman, M.T., and Churchland, A.K. (2014). A category-free neural population supports evolving demands during decision-making. Nature Neuroscience *17*, 1784-1792.
Rathelot, J.-A., and Strick, P.L. (2009). Subdivisions of primary motor cortex based on cortico-motoneuronal cells. Proceedings of the National Academy of Sciences of the United States of America *106*, 918-923.
Regan, J.M., Worley, E., Shelburne, C., Pullarkat, R., and Watson, J.C. (2015). Burr Hole Washout versus Craniotomy for Chronic Subdural Hematoma: Patient Outcome and Cost Analysis. PloS one *10*, e0115085.
Revechkis, B., Aflalo, T.N., Kellis, S., Pouratian, N., and Andersen, R.A. (2014). Parietal neural prosthetic control of a computer cursor in a graphical-user-interface task. J Neural Eng *11*, 066014.
Rigotti, M., Barak, O., Warden, M.R., Wang, X.J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585-590.
Rizzolatti, G., and Fabbri-Destro, M. (2008). The mirror system and its role in social cognition. Current Opinion in Neurobiology *18*, 179-184.
Rozzi, S., Ferrari, P.F., Bonini, L., Rizzolatti, G., and Fogassi, L. (2008). Functional organization of inferior parietal lobule convexity in the macaque monkey: electrophysiological characterization of motor, sensory and mirror responses and their correlation with cytoarchitectonic areas. European Journal of Neuroscience *28*, 1569-1588.
Salakhutdinov, R., Tenenbaum, J.B., and Torralba, A. (2013). Learning with hierarchical-deep models. IEEE transactions on pattern analysis and machine intelligence *35*, 1958-1971.
Salinas, E., and Sejnowski, T.J. (2001). Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet. The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry *7*, 430-440.
Salinas, E., and Thier, P. (2000). Gain modulation: a major computational principle of the central nervous system. Neuron *27*, 15-21.
Schaffelhofer, S., Agudelo-Toro, A., and Scherberger, H. (2015). Decoding a wide range of hand configurations from macaque motor, premotor, and parietal cortices. The Journal of Neuroscience *35*, 1068-1081.
Schaffelhofer, S., and Scherberger, H. (2016). Object vision to hand action in macaque parietal, premotor, and motor cortices. Elife *5*, e15278.
Schalk, G., Kubanek, J., Miller, K.J., Anderson, N.R., Leuthardt, E.C., Ojemann, J.G., Limbrick, D., Moran, D., Gerhardt, L.A., and Wolpaw, J.R. (2007). Decoding two-dimensional movement trajectories using electrocorticographic signals in humans. J Neural Eng *4*, 264-275.
Sergio, L., Hamel-Paquet, C., and Kalaska, J. (2005). Motor cortex neural correlates of output kinematics and kinetics during isometric-force and arm-reaching tasks. Journal of neurophysiology *94*, 2353-2378.
Shamir, M., and Sompolinsky, H. (2006). Implications of neuronal diversity on population coding. Neural computation *18*, 1951-1986.
Silla, C.N., and Freitas, A.A. (2011). A survey of hierarchical classification across different application domains. Data Mining and Knowledge Discovery *22*, 31-72.

Simeral, J.D., Kim, S.-P., Black, M.J., Donoghue, J.P., and Hochberg, L.R. (2011). Neural control of cursor trajectory and click by a human with tetraplegia 1000 days after implant of an intracortical microelectrode array. Journal Of Neural Engineering *8*, 025027.
Singh, C., and Levy, W.B. (2017). A consensus layer V pyramidal neuron can sustain interpulse-interval coding. PloS one *12*, e0180839.
Sliwa, J., Planté, A., Duhamel, J.-R., and Wirth, S. (2016). Independent Neuronal Representation of Facial and Vocal Identity in the Monkey Hippocampus and Inferotemporal Cortex. Cerebral Cortex *26*, 950-966.
Snyder, L.H., Batista, A.P., and Andersen, R.A. (1997). Coding of intention in the posterior parietal cortex. Nature *386*, 167-170.
Snyder, L.H., Grieve, K.L., Brotchie, P., and Andersen, R.A. (1998). Separate body- and world-referenced representations of visual space in parietal cortex. Nature *394*, 884-891.
Stein, R.B., Gossen, E.R., and Jones, K.E. (2005). Neuronal variability: noise or part of the signal? Nature Reviews Neuroscience *6*, 389-397.
Sugarman, M.A., Woodard, J.L., Nielson, K.A., Seidenberg, M., Smith, J.C., Durgerian, S., and Rao, S.M. (2012). Functional Magnetic Resonance Imaging of Semantic Memory as a Presymptomatic Biomarker of Alzheimer's Disease Risk. Biochimica et biophysica acta *1822*, 442-456.
Tai, Y.F., Scherfler, C., Brooks, D.J., Sawamoto, N., and Castiello, U. (2004). The human premotor cortex is 'mirror' only for biological actions. Curr Biol *14*, 117-120.
Tangney, J.P., Stuewig, J., and Mashek, D.J. (2007). Moral Emotions and Moral Behavior. Annual Review of Psychology *58*, 345-372.
Taylor, D.M., Tillery, S.I.H., and Schwartz, A.B. (2002). Direct cortical control of 3D neuroprosthetic devices. Science *296*, 1829-1832.
Thrun, S., and O'Sullivan, J. (1998). Clustering Learning Tasks and the Selective Cross-Task Transfer of Knowledge. In Learning to Learn, S. Thrun, and L. Pratt, eds. (Boston, MA: Springer US), pp. 235-257.
Tibshirani, R., Ealther, G., and Hastie, T. (2001). Estimating the number of clusters in a dataset via the Gap statistic. JR Statist Soc *63*, 411-423.
Tsao, D.Y., Freiwald, W.A., Tootell, R.B.H., and Livingstone, M.S. (2006). A Cortical Region Consisting Entirely of Face-Selective Cells. Science *311*, 670-674.
Ungerleider, L.G., and Mishkin, M. (1982). Two cortical visual systems. In Analysis of Visual Behavior, D.J. Ingle, M.A. Goodale, and R.J.W. Mansfield, eds. (Cambridge, MA: MIT Press), pp. 549-585.
Waal, F.B.M.d. (2008). Putting the Altruism Back into Altruism: The Evolution of Empathy. Annual Review of Psychology *59*, 279-300.
Wurm, M.F., and Lingnau, A. (2015). Decoding Actions at Different Levels of Abstraction. The Journal of Neuroscience *35*, 7727-7735.
Yosinski, J., Clune, J., Bengio, Y., and Lipson, H. (2014). How transferable are features in deep neural networks? Advances in Neural Information Processing Systems *27*, 3320-3328.
Yttri, E.A., Wang, C., Liu, Y., and Snyder, L.H. (2014). The parietal reach region is limb specific and not involved in eye-hand coordination. J Neurophysiol *111*, 520-532.
Zhang, C.Y., Aflalo, T., Revechkis, B., Rosario, E.R., Ouellette, D., Pouratian, N., and Andersen, R.A. (2017). Partially Mixed Selectivity in Human Posterior Parietal Association Cortex. Neuron *95*, 697-708.
Zhang, Y., Meyers, E.M., Bichot, N.P., Serre, T., Poggio, T.A., and Desimone, R. (2011). Object decoding with attention in inferior temporal cortex. Proceedings of the National Academy of Sciences *108*, 8850-8855.

Zhou, J., Chen, J., and Ye, J. (2011). Clustered Multi-Task Learning Via Alternating Structure Optimization. Adv Neural Inf Process Syst *2011*, 702-710.
Zipser, D., and Andersen, R.A. (1988). A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature *331*, 679-684.

8 Appendix

8.1 Supplementary Data

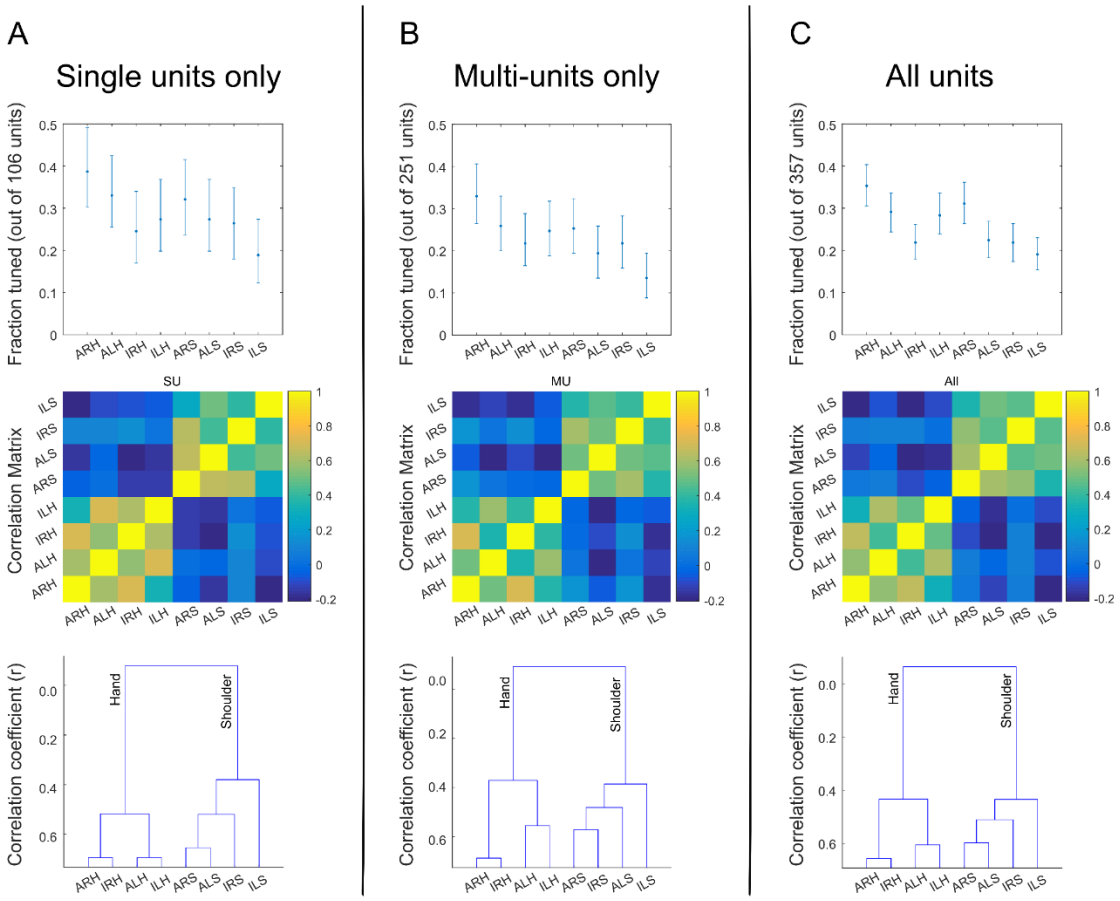

Figure S 1. Using only high-quality single units or multi-units does not qualitatively change results

(A) Top: Fraction of units significantly tuned to each movement condition when considering only high-quality single units ($p < 0.05$, uncorrected 95% confidence intervals). Middle: Pairwise correlation matrix between conditions for high-quality single units. Color scale normalized to span the 5th to 95th percentiles of the correlation values. Bottom: Dendrogram showing hierarchical organization of responses to the different conditions (using correlation as a distance measure), considering only high-quality single units. (B) Similar to (A) but considering more poorly isolated units. (C) Similar to (A) but considering all units (both single units and multi-units). (ALH = Attempt Left Hand, ILH = Imagine Left Hand, ARH = Attempt Right Hand, IRH = Imagine Right Hand, ALS = Attempt Left Shoulder, ILS = Imagine Left Shoulder, ARS = Attempt Right Shoulder, IRS = Imagine Right Shoulder).

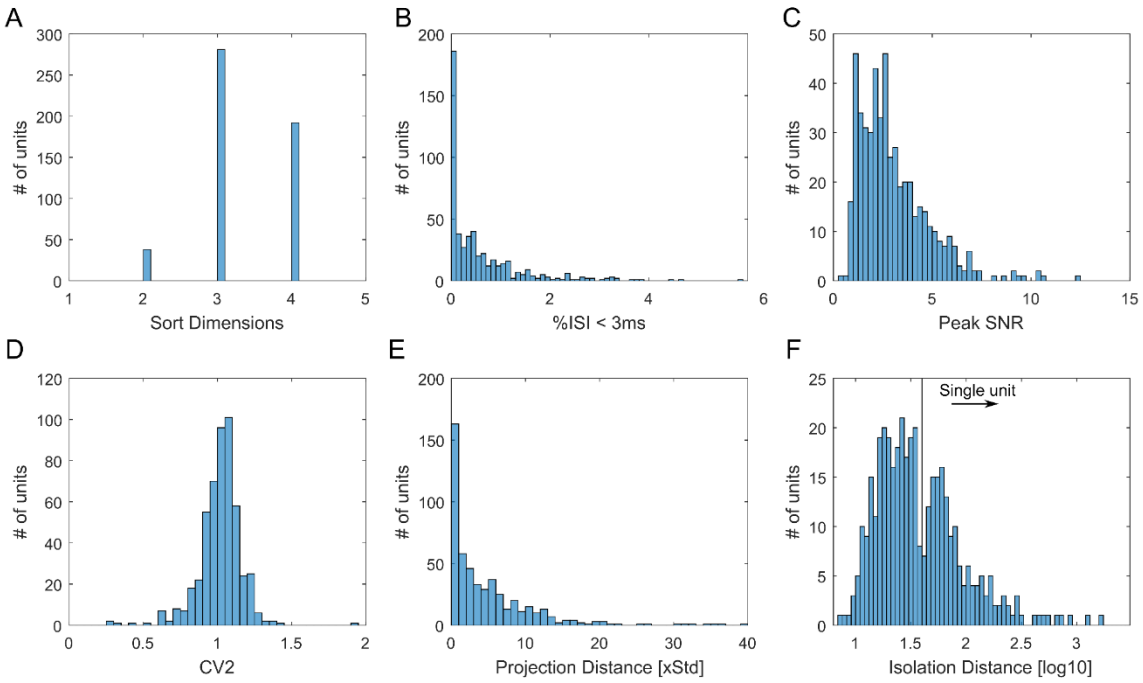

Figure S 2. Metrics of single unit cluster isolation quality

(A) Distribution of the number of principal components used in the clustering algorithm to isolate each unit (i.e., the number of principal components making up 95% of the variance for each unit). (B) Distribution of the percentage of ISIs less than 3ms for each unit. (C) Distribution of the peak signal-to-noise ratio for each unit. (D) Distribution of the modified coefficient of variation of the ISI for each unit. (E) Distribution of the average between-spike projection distance for each unit. (F) Distribution of the base-10 log of the cluster isolation distances for each unit. The isolation distance threshold used to identify high-quality single units (see Unit quality classification in STAR Methods) is indicated by the vertical line.

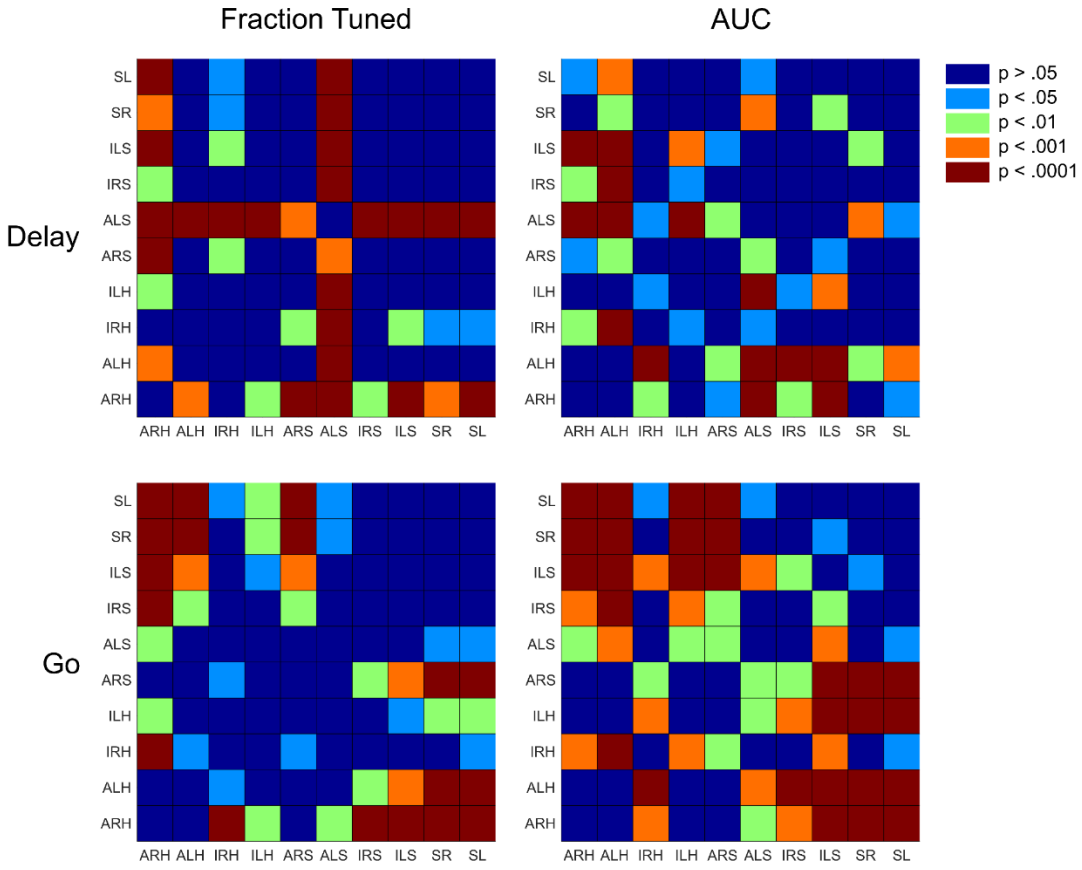

Figure S 3. P-values from two-sided Wilcoxon rank sum tests of whether the fraction of the population tuned to each condition or the AUC values for each condition are significantly different

Figure: P-values from Wilcoxon rank sum tests of whether the conditions' fraction tuned and AUC are significantly different. Left side: Tests done using fraction tuned to each condition. Right side: Tests done using AUC values for each condition. Top row: Tests done using data from the Delay phase. Bottom row: Tests done using data form the Go phase. Significance values are color coded with red being significant with $p < 0.0001$ and dark blue being insignificant ($p > 0.05$). (ALH = Attempt Left Hand, ILH = Imagine Left Hand, ARH = Attempt Right Hand, IRH = Imagine Right Hand, ALS = Attempt Left Shoulder, ILS = Imagine Left Shoulder, ARS = Attempt Right Shoulder, IRS = Imagine Right Shoulder, SR = Speak Right, SL = Speak Left).

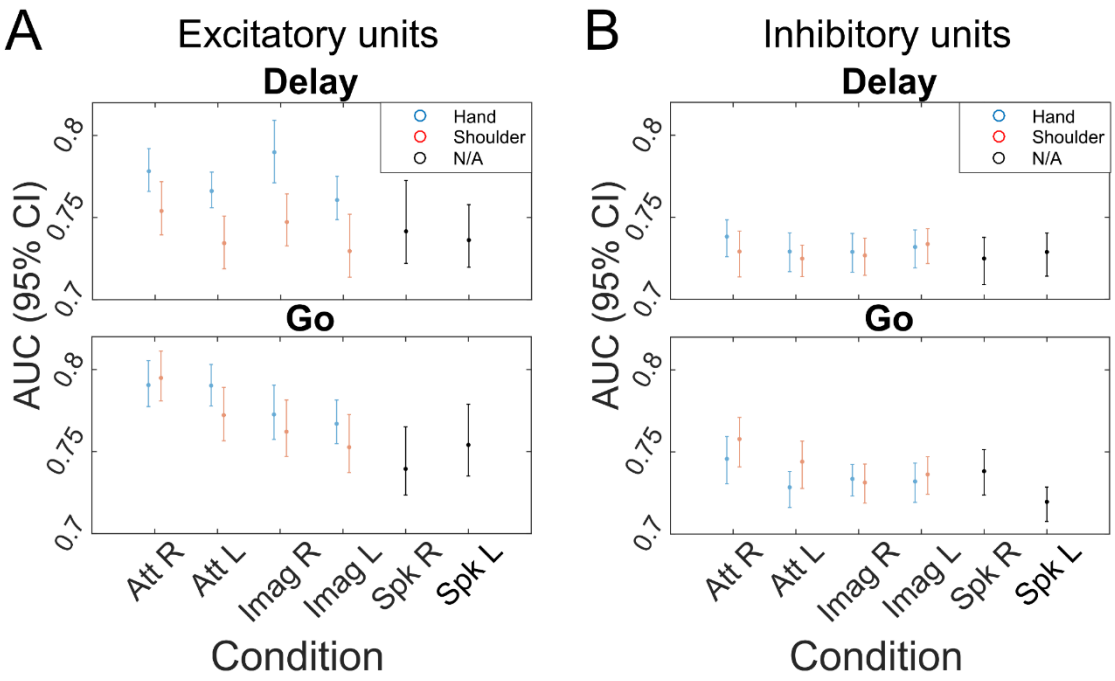

Figure S 4. AUC of units for each condition is comparable between excitatory (positively tuned) and inhibitory (negatively tuned) units

(A) AUC values for excitatory units, split by movement condition (strategy, body side, and body part). Error bars represent the 95% bootstrapped confidence intervals of the AUC values. Top: AUC for units during the Delay phase. Bottom: AUC for units during the Go phase. (B) Similar to (A) but for only inhibitory units. Condition labels were flipped for presentation of AUC values for the inhibitory units for ease of comparison (see STAR Methods). (Att R = Attempt Right, Att L = Attempt Left, Imag R = Imagine Right, Imag L = Imagine Left, Spk R = Speak Right, Spk L = Speak Left).

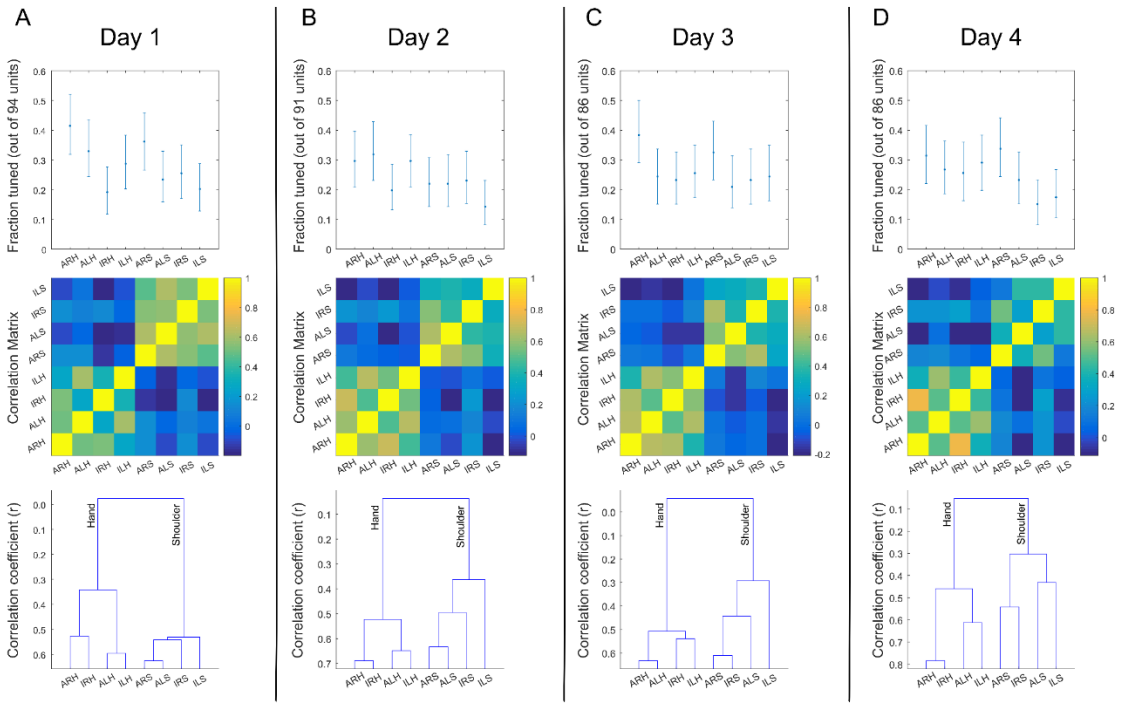

Figure S 5. Results are consistent across separate days of recording sessions

(A) Top: Fraction of units significantly tuned to each movement condition for Day 1 ($p < 0.05$, uncorrected 95% confidence intervals). Middle: Pairwise correlation matrix between movement conditions for Day 1. Color scale normalized to span the 5th to 95th percentiles of the correlation values. Bottom: Dendrogram showing hierarchical organization of responses to the different conditions for Day 1. (B-D) Similar to (A) but for Days 2-4, respectively. (ALH = Attempt Left Hand, ILH = Imagine Left Hand, ARH = Attempt Right Hand, IRH = Imagine Right Hand, ALS = Attempt Left Shoulder, ILS = Imagine Left Shoulder, ARS = Attempt Right Shoulder, IRS = Imagine Right Shoulder).

8.2 Supplementary Videos

8.2.1 Sensory Mirroring Video

Example of a sensory mirroring neuron, showing its response when feeling/observing touches on the inside/outside of the shoulder. The neuron prefers *felt* touches to the outside of the shoulder (has a higher firing rate) over felt touches to the inside of the shoulder, near the neck. Likewise, the neuron prefers *observed* touches to the outside of the shoulder over observed touches to the inside of the shoulder. In other words, the representation of the body parts (outside and inside of the shoulder) is consistent regardless of whether the sensation is observed or felt. The response of the neuron can be observed through the audio (with the crackles indicating the neuron firing/being active) as well as through the trace of the neural activity (the pink traces on the monitor in the video).