**SPECIFIC AIMS**

Pursuing goals in a complex environment requires behavioral flexibility – the ability to rapidly adjust actions and expectations in response to changes in internal and environmental variables. Deficits in flexible behavior are seen in nearly all psychiatric conditions, including addiction, anxiety, mood, and psychotic disorders. However, the neural basis of behavioral flexibility remains poorly understood. This proposal will study two important types of flexibility: **Type 1**, rapid context-dependent adjustments in behavior, and **Type 2**, rapid decision-making in novel conditions through generalization from past experiences. These types of flexibility make distinct demands on neural representations [1–4]. The dorsolateral prefrontal cortex (DLPFC) [5–9] and hippocampus (HPC) [1, 10–14] are thought to contribute to both types of flexibility. This proposal seeks to delineate each area’s distinct computational role in these two types of flexible behavior.

Recent work has suggested that analyzing the *geometry* of neural representations can illuminate the computational role of a neural ensemble in supporting both types of flexibility [15, 16]. Geometry here refers to the structured relationships among firing patterns in a neural ensemble that are elicited by different conditions. In the context of experimental data, these relationships can be visualized by plotting the firing rate of neurons for each of experimental conditions The -dimensions of the plot correspond to the firing-rate of each neuron, and each of points corresponds to the firing pattern elicited by an experimental condition. Each experimental condition is defined by the values of the task variables. Recent work suggests that **Type 1** flexibility requires a geometry in which a linear decoder can read out the value of as many variables as possible. This is because different variables acquire behavioral significance depending on task demands, which are often context-dependent. Flexible context-dependent adjustments in behavior will fail if too few variables are represented. **Type 2** flexibility demands a geometry that has certain generalization properties. Specifically, as will be explained below, to support generalization in a novel condition, the geometry should allow a linear decoder to read out variables in novel conditions *in the same way* it reads out variables for familiar conditions. Two measures of the geometry of representations have been proposed to support these types of flexibility [15]. The first, shattering-dimensionality (SD), quantifies how many variables can be read out from a neural representation. The second, cross-condition generalization performance (CCGP), quantifies how well a linear decoder trained to read out a variable from *some* conditions can accurately read out the same variable from conditions not used for training. HPC and DLPFC both contain representations that simultaneously exhibit high SD and high CCGP [15]. However, a relationship between these measures and these two types of flexible behavior remains to be established. I propose to **investigate the behavioral relevance of representational geometry** in HPC and DLPFC as monkeys perform a task requiring both types of behavioral flexibility on separate trials. This proposal will test the **hypothesis** that Type 1 flexibility correlates with SD, and Type 2 flexibility correlates with CCGP.

**Aim 1: To determine if and where shattering-dimensionality (SD) correlates with context-dependent behavioral flexibility.** Monkeys will perform a serial reversal-learning task in which they switch between two contexts, each defined by four experimental conditions. Each experimental condition corresponds to a unique, context-dependent stimulus-response-outcome (SRO) mapping for one of the four visual stimuli used. On the first 8 trials of each block, contextual cues will appear after presentation of the visual stimulus and a brief delay period. Animals will perform many trials in one context before being presented with a switch cue, signaling 50% probability of a context switch on the next trial. Neural responses will be analyzed in the following trial during the delay period that precedes the contextual cue. Since context is uncertain during this period, animals will need to represent all 8 SRO mappings in order to select the correct response after being informed of the context, an example of Type 1 flexibility. As these trials rely on working memory and task switching for the 8 SRO mappings, I hypothesize that the correlation between SD and behavior will be strongest in DLPFC as compared to HPC.

**Aim 2: To determine if and where cross-condition generalization performance (CCGP) correlates with successful generalization in novel conditions.** After performing several blocks of the task in Aim 1, two new contextual cues will be introduced for the two behavioral contexts (again, on the first 8 trials of each block). However, now animals will only be exposed to conditions using three of the four visual stimuli. After experiencing many trials in both contexts, the fourth stimulus will appear for the first time on a trial following a switch cue. Because animals have never experienced the new contextual cues with this stimulus, such trials function as novel conditions. Monkeys can exhibit generalization on this trial by remembering from prior experience that the fourth stimulus is linked to the other three within a context and acting appropriately, an example of Type 2 flexibility. As in Aim 1, neural responses will be analyzed during the delay period following stimulus presentation. Because of the HPCs role in pattern completion and contextual associations and the DLPFCs role in working memory and rule-related processing, I hypothesize that performance on these trials will correlate with CCGP in both HPC and DLPFC.

**RESEARCH STRATEGY**

1. **SIGNIFICANCE:**

*Overview:* We live in a world that is constantly changing. In order to pursue goals and survive in such a complex environment, one cannot rely on learned stimulus-response associations alone. Rather, we require behavioral flexibility – the ability to rapidly adjust actions and expectations in response to changes in internal and environmental variables. Our ability to engage in flexible behavior breaks down under conditions of stress and in the setting of many neuropsychiatric conditions, including addiction, anxiety, mood, and psychotic disorders [17–21]. This project aims to characterize the underlying neurophysiology that gives rise to two important types of behavioral flexibility: **Type 1**, rapid context-dependent adjustments in behavior, and **Type 2**, rapid decision-making in novel conditions through generalization from past experiences. These types of flexibility make distinct demands on neural representations [1–4]. The dorsolateral prefrontal cortex (DLPFC) [5–9] and hippocampus (HPC) [1, 10–14] are thought to contribute to both types of flexibility, but delineating each area’s unique computational role has proved difficult. The overarching hypothesis of this grant is that neural representations of variables – considered as the pattern of activity in a population across experimental conditions - must exhibit a particular *geometry* to support the two types of behavioral flexibility. I will test this hypothesis by comparing the geometry of representations during correct and incorrect behavior on certain trial types. By elucidating the neural basis of flexibility in HPC and DLPFC, this project provides a foundation for understanding the neural basis of deficits seen in stress and neuropsychiatric illness, paving the way for the development of future treatments.

*Neural Representations in Systems Neuroscience:* Neural representations can be considered to be the patterns of activity in across a neural ensemble observed in relation to modulations in internal and environmental variables. Neurons can represent variables that describe “explicit” features of the world (e.g. physical properties or internal states like thirst), and “hidden” (or latent) features, such as relationships between objects or other features [22]. Traditionally, research in neuroscience has sought to measure the *content* of neural representations in efforts to explain complex behaviors. In this way, researchers have identified many important variables that are involved in computations like object recognition, category formation, and rule-based decision-making [23–26]. Efforts of this kind typically begin by postulating that some variable is necessary for the behavior in question and proceed by searching for representations of that variable in candidate brain areas, during key task epochs. However, such studies have typically not distinguished between different types of flexible behavior and how these types place distinct demands in how a neural ensemble represents specific variables to support behavioral flexibility.

*The Geometry of Neural Representations:*Chart, radar chart

Description automatically generated In contrast to this traditional approach, recent work has suggested that analyzing the *geometry* of neural representations can illuminate the computational role of a neural ensemble in supporting Type 1 and Type 2 flexibility [15, 16]. Geometry here refers to the structured relationships among firing patterns in a neural ensemble that are elicited by different conditions. In the context of experimental data, these relationships can be visualized by plotting the firing rate of neurons for each of experimental conditions The -dimensions of the plot correspond to the firing-rate of each neuron, and each of the points correspond to the firing pattern elicited by an experimental condition. Each experimental condition is defined by the values of the task variables, such as stimulus identity, operant response, and reward contingencies. The plots in Fig. 1 schematize two different geometries for 8 experimental conditions across three neurons. Based on each neuron’s selectivity for the variables that comprise each condition, the 8 experimental conditions appear as points in specific regions of the firing-rate space. The center of each point represents the average activity across many trials under that experimental condition. In this figure, the responses of 3 neurons are plotted against each other for each representation, but the intuition extends to higher neuron counts as well.

**Fig. 1: Behavioral flexibility and the geometry of neural representations.** (left) Targeted brain areas in the rhesus macaque. (right) Schematic showing predicted geometries for the two types of behavioral flexibility. Plots show the response of three neurons to 8 experimental conditions belonging to two contexts (blue & green). The top shows a geometry in which all conditions (and variables) are highly separable by a linear decoder, as the representation has high SD. The bottom shows a geometry in which conditions are highly structured across the two contexts, yielding high CCGP and somewhat lower SD.

Recent research from the Salzman and Fusi labs has suggested that **Type 1** flexibility requires a geometry in which a linear decoder can read out the value of as many variables as possible. This is because different variables acquire behavioral significance depending on task demands, which are often context-dependent. Flexible context-dependent adjustments in behavior will fail if too few variables are represented, as the proper input-output mappings cannot be generated if the relevant variables that define such mappings are not present in neural activity. Geometry of this kind is schematized in the upper plot in Fig. 1, where each experimental condition randomly occupies a separate region of the firing-rate space. In fact, because of the arrangement of the points, assuming noise is not too high, a linear decoder can separate all possible groupings of 4 conditions from the remaining 4 conditions. Thus any variable can be read out in relation to task demands.

In contrast, **Type 2** flexibility demands a geometry that has certain generalization properties. Specifically, to support generalization in a novel condition, the geometry should allow a linear decoder to read out variables in novel conditions *in the same way* it reads out variables for familiar conditions. This can occur because the world is structured, and novel conditions often share features with familiar conditions. If the brain represents such structured relationships between conditions, it facilitates generalization, as neural responses to novel conditions can use the brain’s pre-existing encoding scheme for the novel condition’s constituent variables. Such a geometry is shown in the lower plot of Fig. 1, where experimental conditions are not only organized by context, but also along axes of symmetry that could correspond to additional task variables.

Chart, bubble chart

Description automatically generated*Two Measures of Geometry:* The Salzman and Fusi labs have developed two tools to measure the geometric properties thought to underlie Type 1 and Type 2 flexibility [15]. The first, called **shattering-dimensionality (SD)**, quantifies how many variables a linear decoder can be read out from a neural representation. This is schematized in Fig. 2, which shows an example representation of four experimental conditions across two neurons. Together, the 7 panels of the figure show all possible binary variables that could be defined across these 4 conditions. Within each panel, each experimental condition is assigned an arbitrary color, corresponding to the value of the binary variable in question. For example, in Fig. 2a, condition is red, while the other conditions are yellow, indicating that the variable in question takes one value for and another value for the remaining conditions. Using linear decoders, a boundary can be generated that optimally segregates conditions with respect to the variable in question. These decoders are trained with data from *all* experimental conditions that differ with respect to this variable (in contrast to the second method, described below). This boundary is shown as a dotted line in Figs. 2 a-f. By contrast, Fig. 2g shows a variable assignment for which no such boundary can be generated. Because linear decoders can successfully read out variables from 6 of the 7 possible assignments, this representation has a SD of 6/7 = 0.86. Again, this example generalizes to higher-dimensional representations with more neurons and experimental conditions.

**Fig. 2: Computing shattering-dimensionality (SD).** Example representation for four experimental conditions across two neurons. Each panel shows one of the 7 arbitrary variables that could be read out from these four experimental conditions (visualized as yellow vs. red). The dotted line in panels (a-f) show how a linear decoder could separate the four conditions according to the variable in question when training on a subset of all conditions. Panel (g) shows a geometry that does not yield linear separability using a decoder of this kind. Therefore, SD for this representation would be 6/7 = 0.86.

Diagram

Description automatically generatedThe second measure of geometry, called **cross-condition generalization performance (CCGP)**, quantifies how well a linear decoder trained to read out a variable from *some* conditions can accurately read out the same variable from conditions not used for training. This is shown in Fig. 3, which again plots the activity of two neurons against each other for four experimental conditions, where two binary variables describe each condition. To calculate CCGP, linear decoders are again trained to segregate conditions, but this time only using data from *some* conditions, called the training set. This again will yield a boundary that maximally separates conditions based on the variable in question. Then, using this pre-trained network, we evaluate how well this boundary can be used to correctly classify conditions in the testing set according to the variable in question. This is performed for every possible assignment of training and testing sets. A boundary line is shown, which is derived from training the decoder on the designated subset of experimental conditions. Performance of this trained decoder to read out the value of the variable on the testing set varies depending upon the training set. CCGP is the average performance across all possible training/testing sets. For Fig. 3a, CCGP is therefore (1 + 1 + 0.5 + 0.5) ÷ 4 = 0.75. By contrast, if a variable is represented by the pattern of activity shown in Fig. 3b, the geometry does not produce high CCGP. Here CCGP is 0 for the depicted variable. CCGP can be calculated for any and all possible groupings of experimental conditions, where each grouping can be considered a variable.

**Fig. 3: Computing cross-condition generalization performance (CCGP).** Two examples showing a representation as in Fig. 2, with four experimental conditions across two neurons. Panel (a) shows one variable assignment (a red/ yellow dichotomy), while panel (b) shows an alternative variable assignment (blue/ green). The dotted line in each panel shows how a linear decoder could separate the two training conditions. We then measure how well how this decoding boundary generalizes to correctly classify the testing conditions. The variable in panel (a) shows successful generalization properties in the first two sub-panels, while it generalizes at chance levels for the second two. The average CCGP for this variable representation is therefore 0.75. By contrast, panel (b) shows no such generalization properties, with an average CCGP of 0.

Of note, both SD and CCGP differ from traditional investigations of neural representations, as 1) these analyses measure the geometry of representations, and 2) these measures not only characterize ‘task-relevant’ variables but all possible variables that can be assigned to the representation. The methods thereby describe the geometry of representations in a completely unbiased manner [15].

*Prior Results and Outstanding Questions:* Previous work in the lab has examined the geometry of representations using data from a serial reversal-learning task and demonstrated that HPC and DLPFC both contain representations that simultaneously exhibit high SD and high CCGP [15]. Importantly, this work confirmed theoretical predictions that variables may not exhibit high CCGP, even when they are decodable using traditional methods.

However, a relationship between these measures and these two types of flexible behavior remains to be established, as previous studies did not require animals to generalize in novel conditions. This project stands to make crucial advancements over our earlier work by introducing a task that requires both types of behavioral flexibility on separate trials. Together with these theoretical tools, this project promises to be the first to demonstrate the role of the geometry of neural representations in behavioral flexibility, potentially revealing a new and important population-level neural coding scheme [3, 27–29]. My **hypothesis** that Type 1 flexibility correlates with SD, and Type 2 flexibility correlates with CCGP in these brain areas is summarized in Fig. 1.

*Potential Impact:* This project will contribute to the field in three important ways. 1) This will be the first study to show specific evidence connecting the geometry of neural representations to the performance of Type 1 and Type 2 behavioral flexibility. Focusing on the *geometry* of representations contrasts with dominant approaches in neuroscience that have only examined the *content* of variables being represented. While my hypothesis posits that these types of behavioral flexibility are inherently dependent on the geometry of neural representations, these are by no means the only behaviors for which geometry might play an essential role. Demonstrating the importance of geometry of representations for complex behavior will open the field to further investigations into coding schemes of this kind. 2) This study represents a unique neurobiological application of the recent advances in machine-learning and theories of neural computation. In machine-learning, generalization is assessed by testing a network’s classification performance using samples that are different in some important respect from the training samples. By employing a symmetric task-design with novel stimuli, I will be able to apply this definition to the analysis of my neural data, training a network on *some* experimental conditions and testing its classification performance on *other* experimental conditions. By correlating this cross-condition generalization performance with behavioral generalization, this project has the potential to reveal an important computational mechanism behind the generation of flexible behavior in novel situations. 3) By performing high channel-count recordings from two different brain areas simultaneously, I will be able to compare the time-dependent correlations between geometric format and behavioral performance in these areas. This will provide a new approach to delineate the function of different brain areas, paving the way for development of future treatments for neuropsychiatric illness.

1. **APPROACH:**

*Overall Methods:* Animals will be trained to perform a task that requires Type 1 and Type 2 behavioral flexibility on separate trials. During behavior, I will perform simultaneous electrophysiological recordings of HPC and DLPFC. HPC will include recordings from across CA1, CA2, CA3, and DG. DLPFC recordings will be taken from Broadman areas 8, 9, and 46. Recordings will be made continuously using 1-2 32-channel vertical electrodes (V-Probes, Plexon) in each brain area. Analog signals will be amplified, filtered, and digitized, after which, each channel will undergo automated spike sorting using MountainSort [30] to separate single units within each channel, which will additionally be verified manually using the Plexon offline sorter platform. Measures of geometry will be performed according to the procedures outlined in the Aims below.

*Sex as a Biological Variable:* In primate electrophysiology studies, studies are typically made in 2 experimental subjects, making it impossible to study sex as a biological variable. Moreover, male monkeys are typically employed because they are more readily available for purchase, with females often reserved for use in breeding colonies. As such, we currently have no female animals in our lab. Because of these limitations, meaningful comparisons of findings across gender will not be possible in the present study.

**Aim 1: To determine if and where shattering-dimensionality (SD) correlates with context-dependent behavioral flexibility.**

*Rationale:* Context-dependent behavior requires keeping track of many variables that describe features of each context. Context-depending behavior should suffer when too few variables are represented. In this Aim, monkeys will perform a task that requires context-dependent behavioral flexibility. I will analyze the shattering-dimensionality (SD) of neural representations in successful and unsuccessful trials to assess how geometries that encode many variables can support this type of flexibility.

*Experimental Procedures:* Two male rhesus macaque monkeys will be trained to perform a serial reversal-learning task in which they switch between two contexts, each defined by four experimental conditions. Each experimental condition corresponds to a unique, context-dependent stimulus-response-outcome (SRO) mapping for one of the four visual stimuli used. The visual stimuli will be computer-generated fractal images (A-D), the operant responses will be one of four saccadic eye movements (up, down, left, right), and the outcome of successful trials will be either a liquid reward or no reward. Each of these SRO mappings is shown in Fig. 4a. This design ensures that any given operant response is not preferentially linked to a particular reinforcement outcome, though fractals B and D remain rewarded and unrewarded across both contexts, respectively. Fig. 4b shows the trial structure. Animals maintain central fixation for 500 ms, followed by stimulus presentation for 300 ms. On the first 8 trials of each block, contextual cues (also fractal images) will appear for 300 ms following a delay period of 1000 ms. For the subsequent trials in a block, no contextual cue will appear. Saccade targets will appear during a second delay period of 1000 ms, and the removal of the central fixation point will signal the animal to respond. Correct responses will be followed by a liquid reward, depending on the SRO mapping for the experimental condition.

Animals will perform many trials in one context before being presented with a switch cue, signaling 50% probability of a context switch on the next trial. In the following trial, context will be maximally uncertain, and animals will need to wait for the contextual cue to know the correct response and outcome. Thus, the task demands that the monkeys represent all possible 8 experimental conditions (SRO mappings) in a distinct manner pending being informed of the context. Only then can the monkey determine the correct response by knowing the stimulus and the context and the associated contingencies. On trials in which Graphical user interface

Description automatically generatedcontext does not change, monkeys will be presented with another switch-cue trial until they enter a new context. This block structure is shown in Fig. 4c. To keep animals engaged, the base set of visual stimuli (A-D) will be replaced after several days of consistently high performance. While animals perform the task, several behavioral measures will be recorded, including anticipatory licking behavior, and continuous eye-movement tracking and pupillometry. Anticipatory licking will be recorded with a custom-built apparatus. Eye movements and pupillometry will be recorded with commercial equipment from SR Systems (EyeLink).

*Preliminary Studies:* This task combines features of tasks previously used in successful studies with behaving monkeys. Namely, it adapts the reversal-learning task in [15], by introducing saccadic eye movements as the operant response, analogous to the design of [31]. However, the contextual and switch cues are novel features of this task design, which serve to create a trace interval in which to analyze the representational geometry preceding behavior. Training has already begun with the first monkey, who is showing evidence of contextual switching using only two of the four fractal images (data not shown).

**Fig. 4: Overall task design.** (a) 8 experimental conditions, defined by four unique stimulus-response-outcome (SRO) mappings across two contexts. (b) Trial structure. Neural activity will be analyzed in the first delay period, following stimulus presentation. The inset shows cues that could appear in context 1 following the first delay period: a contextual cue on the first 8 trials in a block, no cue on subsequent trials, or a switch cue at the end of a block. (c) Block structure, showing transitions between contexts.

*Expected Outcomes:* Neural responses will be primarily analyzed in trials following a switch-cue trial, during the delay period that precedes the contextual cue. Since context is uncertain during this period, animals will need to represent all 8 SRO mappings in order to select the correct response after being informed of the context, an example of Type 1 flexibility. I will compute SD separately using data from successful and unsuccessful trials to correlate this measure of geometry with performance. As these trials rely on working memory and task switching, known functions of DLPFC, I hypothesize that the correlation between SD and behavior will be strongest in DLPFC as compared to HPC.

*Potential Pitfalls:* As with all animal studies, the possibility of multiple behavioral tactics presents the greatest potential pitfall. Namely, there is some concern that animals could select an action on seeing the stimulus, and then switch the action if they are informed of a change in context. However, this tactic is highly unlikely in the current task design, as each stimulus is associated with two different target directions. If the task had used a binary operant response as in [15], this tactic would have posed a greater concern.

Another possible pitfall could arise if animals perform too well on the task, thereby decreasing the number of error trials available for analysis. Should this situation occur, it can be remedied through small changes in the task parameters. For instance, animals can be introduced to a new base set of visual stimuli more frequently to increase the difficulty of the task.

**Aim 2: To determine if and where cross-condition generalization performance (CCGP) correlates with successful generalization in novel conditions.**

*Rationale:* Generalization from past conditions to novel ones requires that the brain represent one or more variables in a manner that can readily classify the novel condition. This can occur if the geometry of a representation already reflects the links between variables comprising familiar (already experienced) and novel conditions. Here I will test whether such geometries correlate with Type 2 flexibility, by comparing the cross-condition generalization performance (CCGP) of neural representations in successful and unsuccessful instances of generalization in novel conditions.

Graphical user interface, Teams

Description automatically generated*Experimental Procedures:* This aim adapts the task in Aim 1, but introduces many trials in which animals must generalize in novel conditions. The block structure for this task shown in Fig. 5. After performing several blocks of the task in Aim 1, two new contextual cues (fractal images) will be introduced for the two behavioral contexts (again, on the first 8 trials of each block). However, now animals will only be exposed to conditions using three of the four visual stimuli. After experiencing many trials in both contexts, the fourth stimulus will appear for the first time on a trial following a switch cue. Because animals have never experienced the new contextual cues with this stimulus, such trials function as novel conditions. Monkeys can exhibit generalization on this trial by remembering from prior experience that the fourth stimulus is linked to the other three within a context and acting appropriately, an example of Type 2 flexibility.

**Fig. 5: Task design for Aim 2.** At first, animals will experience several blocks in each context with all four visual stimuli (A-D). Then new contextual cues will be introduced, and animals will only experience three of the four visual stimuli (A-C, in the first example), before experiencing the held-out stimulus for the first time on a trial following a switch-cue trial. Importantly, as in Aim 1, switch-cue trials will precede all changes in context, with the switch cue signaling a 50% chance of changing context in the following trial. The example here shows the held-out stimulus (D) appearing on a trial when context remains the same.

*Expected Outcomes:* As in Aim 1, neural responses will be primarily analyzed during the delay period following stimulus presentation. I will compute CCGP separately for the variable context using data from successful and unsuccessful trials to correlate this measure of geometry with Type 2 flexibility. To do so, I will analyze activity from the delay period following switch-cue trials, which will include data from the three previously experienced conditions as well as the held-out (novel) condition, as indicated by the dotted line in the figure. Across both contexts, this will yield delay period activity from all 8 experimental conditions. The HPC has long been implicated in the formation of episodic memories [32–34], which due to their relational nature, might be involved in the creation and maintenance of representations supporting generalization. Human studies suggest that the HPC is involved in conceptual knowledge and learning of statistical regularities across events [35, 36]. In single-neuron experiments, DLPFC has also been shown to encode rule and category-based information in working memory [24, 25, 37–39]. Furthermore, both of these areas have been implicated in representing variables related to both novelty and task structure [26, 40–42]. Against this background, I hypothesize that performance on these trials will correlate with CCGP in both HPC and DLPFC. Note that this version of the task, as in Aim 1, employs a switch cue signaling a 50% chance of changing context on the next trial. Consequently, since the context in effect is not indicated until the end of the delay period on the following trial, there is maximal uncertainty as to which SRO mapping might be needed. SD should therefore remain high to support task performance here as well. This implies that neural ensembles may realize a geometry with *both* high CCGP for context and high SD, as was observed in [15]. I therefore predict that performance on trials following a switch cue should also be correlated with SD, as monkeys must link the held-out stimulus to the correct context for generalization *and* must represent the 8 SRO mappings for context-dependent flexibility.

*Potential Pitfalls:* One potential pitfall is that I will fail to see the predicted trends in failure trials due to changes in geometry across the learning process (i.e., as animals *learn to generalize*). It is conceivable that earlier trials are marked by lower CCGP scores in both correct and error trials, which might diminish any comparisons made between the two outcomes. Two approaches are possible in mitigating this confound. First, I will conduct separate analyses by learning epoch, to show that CCGP score increases with overall performance on correct trials but fails to do so with error trials. Also, I will restrict my analysis to trials made only after a criterion level of performance is reached, thereby controlling for the learning process.

**BIBLIOGRAPHY**

[1] Rudy JW, O’Reilly RC. Contextual fear conditioning, conjunctive representations, pattern completion, and the hippocampus. *Behav Neurosci* 19991201; 113: 867.

[2] Fusi S, Miller EK, Rigotti M. Why neurons mix: high dimensionality for higher cognition. *Curr Opin Neurobiol* 2016; 37: 66–74.

[3] Rigotti M, Barak O, Warden MR, et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* 2013; 497: 585--590.

[4] Badre D, Bhandari A, Keglovits H, et al. The dimensionality of neural representations for control. *Curr Opin Behav Sci* 2021; 38: 20–28.

[5] Wallis JD, Anderson KC, Miller EK. Single neurons in prefrontal cortex encode abstract rules. *Nature* 2001; 411: 953--956.

[6] Saez A, Rigotti M, Ostojic S, et al. Abstract context representations in primate amygdala and prefrontal cortex. *Neuron* 2015; 87: 869--881.

[7] Antzoulatos EG, Miller EK. Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. *Neuron* 2011; 71: 243--249.

[8] Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 2001; 24: 167–202.

[9] Mante V, Sussillo D, Shenoy KV, et al. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* 2013; 503: 78–84.

[10] Constantinescu AO, O’Reilly JX, Behrens TE. Organizing conceptual knowledge in humans with a gridlike code. *Science* 2016; 352: 1464--1468.

[11] Sakon JJ, Suzuki WA. A neural signature of pattern separation in the monkey hippocampus. *Proc Natl Acad Sci* 2019; 116: 9634--9643.

[12] Whittington JCR, Muller TH, Mark S, et al. The Tolman-Eichenbaum Machine: Unifying Space and Relational Memory through Generalization in the Hippocampal Formation. *Cell* 2020; 183: 1249-1263.e23.

[13] Gulli RA, Duong LR, Corrigan BW, et al. Context-dependent representations of objects and space in the primate hippocampus during virtual navigation. *Nat Neurosci* 2020; 23: 103–112.

[14] Rolls E. The mechanisms for pattern completion and pattern separation in the hippocampus. *Front Syst Neurosci*; 7. Epub ahead of print 2013. DOI: 10.3389/fnsys.2013.00074.

[15] Bernardi S, Benna MK, Rigotti M, et al. The Geometry of Abstraction in the Hippocampus and Prefrontal Cortex. *Cell* 2020; 183: 954-967.e21.

[16] Kriegeskorte N, Kievit RA. Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci* 2013; 17: 401–412.

[17] Arnsten AFT. Stress weakens prefrontal networks: molecular insults to higher cognition. *Nat Neurosci* 2015; 18: 1376–1385.

[18] Benzina N, Mallet L, Burguière E, et al. Cognitive Dysfunction in Obsessive-Compulsive Disorder. *Curr Psychiatry Rep* 2016; 18: 80.

[19] Koob GF, Volkow ND. Neurocircuitry of Addiction. *Neuropsychopharmacology* 2010; 35: 217–238.

[20] Waltz JA. The neural underpinnings of cognitive flexibility and their disruption in psychotic illness. *Neuroscience* 2017; 345: 203–217.

[21] Park J, Moghaddam B. Impact of anxiety on prefrontal cortex encoding of cognitive flexibility. *Neuroscience* 2017; 345: 193–202.

[22] Behrens TEJ, Muller TH, Whittington JCR, et al. What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* 2018; 100: 490–509.

[23] DiCarlo JJ, Zoccolan D, Rust NC. How Does the Brain Solve Visual Object Recognition? *Neuron* 2012; 73: 415–434.

[24] Wallis JD, Anderson KC, Miller EK. Single neurons in prefrontal cortex encode abstract rules. *Nature* 2001; 411: 953–956.

[25] Miller EK, Nieder A, Freedman DJ, et al. Neural correlates of categories and concepts. *Curr Opin Neurobiol* 2003; 13: 198–203.

[26] Saez A, Rigotti M, Ostojic S, et al. Abstract Context Representations in Primate Amygdala and Prefrontal Cortex. *Neuron* 2015; 87: 869–881.

[27] Fusi S, Miller EK, Rigotti M. Why neurons mix: high dimensionality for higher cognition. *Curr Opin Neurobiol* 2016; 37: 66–74.

[28] Barak O, Rigotti M, Fusi S. The Sparseness of Mixed Selectivity Neurons Controls the Generalization–Discrimination Trade-Off. *J Neurosci* 2013; 33: 3844–3856.

[29] Johnston WJ, Palmer SE, Freedman DJ. Nonlinear mixed selectivity supports reliable neural computation. *PLOS Comput Biol* 2020; 16: e1007544.

[30] Chung JE, Magland JF, Barnett AH, et al. A Fully Automated Approach to Spike Sorting. *Neuron* 2017; 95: 1381-1394.e6.

[31] Fusi S, Asaad WF, Miller EK, et al. A Neural Circuit Model of Flexible Sensorimotor Mapping: Learning and Forgetting on Multiple Timescales. *Neuron* 2007; 54: 319–333.

[32] Milner B, Squire LR, Kandel ER. Cognitive Neuroscience and the Study of Memory. *Neuron* 1998; 20: 445–468.

[33] Eichenbaum H. Hippocampus: Cognitive Processes and Neural Representations that Underlie Declarative Memory. *Neuron* 2004; 44: 109–120.

[34] Wirth S, Yanike M, Frank LM, et al. Single Neurons in the Monkey Hippocampus and Learning of New Associations. *Science* 2003; 300: 1578–1581.

[35] Schapiro AC, Turk‐Browne NB, Norman KA, et al. Statistical learning of temporal community structure in the hippocampus. *Hippocampus* 2016; 26: 3–8.

[36] Kumaran D, Summerfield JJ, Hassabis D, et al. Tracking the Emergence of Conceptual Knowledge during Human Decision Making. *Neuron* 2009; 63: 889–901.

[37] Buckley MJ, Mansouri FA, Hoda H, et al. Dissociable Components of Rule-Guided Behavior Depend on Distinct Medial and Prefrontal Regions. *Science* 2009; 325: 52–58.

[38] Antzoulatos EG, Miller EK. Differences between Neural Activity in Prefrontal Cortex and Striatum during Learning of Novel Abstract Categories. *Neuron* 2011; 71: 243–249.

[39] Wutz A, Loonis R, Roy JE, et al. Different Levels of Category Abstraction by Different Dynamics in Different Prefrontal Areas. *Neuron* 2018; 97: 716-726.e8.

[40] Brown MW, Aggleton JP. Recognition memory: What are the roles of the perirhinal cortex and hippocampus? *Nat Rev Neurosci* 2001; 2: 51–61.

[41] Jutras MJ, Buffalo EA. Recognition memory signals in the macaque hippocampus. *Proc Natl Acad Sci* 2010; 107: 401–406.

[42] Ninokura Y, Mushiake H, Tanji J. Integration of Temporal Order and Object Information in the Monkey Lateral Prefrontal Cortex. *J Neurophysiol* 2004; 91: 555–560.