

Technische Universität Berlin
Fakultät IV – Elektrotechnik und Informatik
Modellierung kognitiver Prozesse

Bachelorarbeit

Impact of Behavioural and Task-Related Variables on Representational Drift in Mouse Visual Cortex

Kira Cordes
Informatik B.Sc.
Matrikel-Nr. 402139

Berlin, 23.05.2023

Betreut von Dr. Felix Lundt und Prof. Dr. Henning Sprekeler

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Berlin, den _____

[Vorname Name]

Zusammenfassung

Bei dem Versuch zu verstehen, wie das Gehirn die Außenwelt intern abbildet, sind Neurowissenschaftlerinnen an den Mechanismen des *neuronalen Codes* interessiert. Selbst im sensorischen Kortex zeigen Hirnareale, wie der primäre visuelle Kortex, die direkt mit dem sensorischen Input verbunden sind und daher als konsistente und stabile Informationsbasis gehalten wurden, Veränderungen neuronaler Reaktionen auf (gleiche) visuelle Stimuli im Laufe der Zeit. Das beschriebene Phänomen wird als Repräsentationsdrift bezeichnet. Der Vergleich von Ähnlichkeiten der neuronalen Repräsentationen im Laufe der Zeit, mit Auswirkungen potenziell störenden Variablen, ermöglicht die Dynamiken neuronaler Verarbeitung und die Faktoren, die den Repräsentationsdrift beeinflussen, besser zu verstehen. Da beobachtet wurde, dass der Verhaltenszustand nicht nur die neuronale Aktivität stark moduliert, sondern auch zu Änderungen neuronaler Reaktionen über die Zeit beiträgt, sollte Repräsentationsdrift nicht nur im Stimulusraum charakterisiert werden. Ähnlich führen auch motivationale Aspekte (zum Beispiel in Form von Aufgabenrelevanz) zur Modulierung neuronaler Aktivität in sensorischen Bereichen und damit potenziell zu Veränderungen in neuronalen Reaktionsmustern. In dieser Arbeit wird der Repräsentationsdrift im visuellen Kortex von Mäusen mit extrazellulären elektrophysiologischen Aufzeichnungen aus einer Datenquelle des *Allen Brain Observatory* untersucht, wobei Faktoren neben den externen visuellen Reizen berücksichtigt werden. Auf Grundlage der Forschung von Sadeh und Clopath, die darauf hindeuten, dass der Verhaltenszustand der Tiere zur Verschiebung von Repräsentationen beiträgt, untersuchen wir diesen Zusammenhang. Darüber hinaus erkunden wir, ob und wie aufgabenbezogene Variablen die Repräsentationsstabilität beeinflussen könnten. Unsere Ergebnisse deuten darauf hin, dass Repräsentationsdrift unterschiedlich ausgeprägt über Zeitskalen, Hierarchien von Gehirnstrukturen, sowie Zelltypen hinweg auftritt und von Parametern außerhalb des Stimulusraums, wie dem Verhaltenszustand beeinflusst wird. Dies entspricht den Erwartungen nach aktuellem Stand der Forschung. Darüber hinaus wird diskutiert welchen Einfluss aufgaben-spezifische Faktoren auf die Stabilität neuronaler Repräsentationen haben könnten. Auswirkungen, die sich in unseren Ergebnissen abzeichnen, könnten auf eine höhere Relevanz von aufgabenbezogenen Variablen, wie Belohnungen hindeuten.

Abstract

When trying to understand how the brain represents the external world internally, neuroscientists are interested in the mechanisms of neural code. Even in sensory cortex brain areas, like the primary visual cortex, directly connected to sensory input and therefore considered to hold consistent and stable information for further computation, recent work found that the neuronal responses to visual stimuli changed over time. The described phenomenon is called representational drift. Comparing the similarity of neural representations over time, along potentially confounding variables, allows to better understand the dynamics of neural processing and the factors that impact representational drift. The behavioural state has been identified to not only modulate neural activity strongly, but also contribute to drift, implying that drift should not be characterised solely in the stimulus space. Correspondingly, motivational aspects (for instance through task-relevance) were found to widely regulate neuronal activity in areas related to sensation and might be affiliated to changes in neuronal response patterns. This thesis analyses representational drift in the visual cortex of mice with extracellular electrophysiology recordings from an Allen Brain Observatory data resource, taking factors other than external visual stimuli into account. Based on research by Sadeh and Clopath suggesting that the behavioural state of animals contributes to drift of representations, we examined this association and further looked at how and if task-related variables impact representational stability. We found that representational drift occurs across timescales, hierarchies of brain structures, cell types and is impacted by parameters outside the stimulus-space, consequently substantiating the current state of research. Our Results suggest further, that task-factors impact stability of neuronal representations considerably, indicating relevance of task-related variables (like rewards) for a contribution to representational drift.

Contents

1	Introduction	1
2	Data	3
2.1	Experimental Setup	3
2.2	Neuropixel Probes	5
2.3	Target Areas, Brain Structures and Cell Types	5
3	Results	8
3.1	Representational Drift in the Visual Cortex	9
3.2	Representational Drift Depends on the Behavioural State	16
3.3	Representational Drift is Impacted by Task-Related Variables	18
4	Discussion	24
4.1	Gradual Changes in Neuronal Response Patterns	24
4.2	Impact of Behavioural Variables	25
4.3	Impact of the Change-Detection Task	25
5	Methods	28
5.1	Data Preprocessing	28
5.2	Data Analysis	29
	Bibliography	31
	picture credits	32
	Abbreviations and Symbols	33
	List of Figures	34

1 Introduction

Neural code is the representation of information in neural activity patterns, making each neuron a channel for processing incoming information and integrating it to produce a response signal. Incoming information, more specifically, is sensory input from the external world and from the internal state; the brain can therefore be seen as an "information-processing machine" (Richmond, 2013). Looking at these internal representations of stimuli, numerous papers and research found, that neuronal responses change over time in various contexts and refer to it as representational drift (Deitch et al., 2021; Driscoll et al., 2022; Driscoll et al., 2017; Rule et al., 2019; Sadeh and Clopath, 2022). Although widely discussed in a substantial body of literature, the underlying mechanisms of ongoing changes in neuronal activity are not well understood. Finding that representations of sensory stimuli are in fact not showing the assumed neural-code stability raised the question of how this can be a basis for coherent interaction with stable long-term representations, persistence of memory, as well as stable sensory perception and motor outputs, or even result into consistent behaviour. Especially lower-level brain structures, closely related to sensory input, have been assumed to exhibit a stable neural-code, but in fact representational drift was found to not be reflected by the hierarchy of information flow across areas (Deitch et al., 2021). In case of the visual cortex, where visual information of eye movements and perception of visual stimuli are integrated and matched with possibly existing experiences, the first stage of receiving and processing sensory information from the eyes takes place in the primary visual cortex (V1). Consequently, higher-order brain or downstream areas are considered to be responsible for associating information. Deitch et al. (2021) further found representational drift across cell types (inhibitory and excitatory) and time-scales. Deviating from the idea of characterizing changes in neural representations solely in the stimulus-space, Sadeh and Clopath (2022) recently looked at the behavioural state of mice, as it modulates neural activity and suggest that behavioural variables, like pupil area and running speed, contribute to drift. The pupil area, for instance, acts as an indicator for arousal of the animal and strongly modulates neuronal activation. Thus, they come to the conclusion that neural activity induced by behaviour could potentially be mistaken as representational drift. To fully explore this notion, it is necessary to take additional factors, like attention and motivation, into account. Both have been shown to modulate responses (Tang et al., 2023) and might therefore be associated with drift. They correspond to task-related variables more accurately as a consequence of task-relevance. Another relevant variable to examine in this context might be task-novelty:

When encountering an unexpected stimulus, the brain generates a significantly larger evoked response compared with the response following an expected stimulus (Garrido et al., 2009, as cited in Tang et al., 2023, p.1)

We therefore analysed neural response data to visual stimuli in mouse visual cortex over time, while performing a change-detection task with particular emphasise on V1 population, as it is the first stage of visual processing. Along with the electrophysiology data we are taking behavioural and task-related variables into account to shed light on the underlying mechanisms of neural processing.

Accordingly, we are splitting our analysis into three main sections, basing the first on Deitch et al. (2021), to examine occurrences of drift in the underlying data, taking aspects like hierarchy or cell types into account. Secondly and based on Sadeh and Clopath (2022)'s work, we are analysing representational drift in relation to behaviour, defined by the animals pupil area and running speed. Finally we are exploring task-environment factors, taking one step further in the third part of our analysis to consider potential impacts of task-novelty or task-relevance on representational drift.

The underlying data originated from a similar experimental-background as the majority of our references and is therefore providing a level of consistency in line with the datasets of Deitch et al. (2021) and Sadeh and Clopath (2022), while inflicting the opportunity to study influences on neural activity of a variation of control variables and task-related factors, from reward volumes, changing visual stimuli to performance of mice in or in absence of the change detection task. The experiment was conducted and recorded by Allen Brain Institute, which will be outlined in the following chapter (2) and is, as a part of Allen Brain Observatory openly accessible. Apart from the extent of dimensions these data resources cover, they are cleaned, preprocessed, eminently consistent with recent work, that we are partly replicating and hence eligible for our research questions. To see whether neural representations drift, we compared neural responses between two points in time and see how similar they are. Decreasing similarity between neural representations to visual stimuli indicates representational drift, which we found across timescales spanning minutes to hours, across brain hierarchies (in VIS and CA1) and in varying intensity. Our results further support recent findings by indicating a correlation between differences in the behavioural state and the similarity of the respective representations. Regarding task-related variables, our analysis implies considerable impacts of the change-detection task on representational drift, although we did not find indications for the influence of stimulus-novelty, while task-relevance and motivation emerge as potentially relevant variables.

2 Data

This chapter gives an overview of the data origins, the experimental setup as well as which and how neuronal responses were recorded. The dataset is part of the openly accessible Allen Brain Observatory data resource collection for understanding sensory processing in the mouse visual cortex and therefore highly eligible for studying representational drift. For our analysis we chose the Visual Behavior Neuropixels¹ (published July 2022) resource, a large-scale and highly standardised electrophysiology recording of mice performing a change-detection task (see fig. 2.1). Additionally to impacts of the behavioural state of mice, the data resource gives us the opportunity to examine associations of representational stability and task-related variables since a visual experiment has been conducted on mice while recording neural activity in various brain regions, along with behavioural and task-related events.

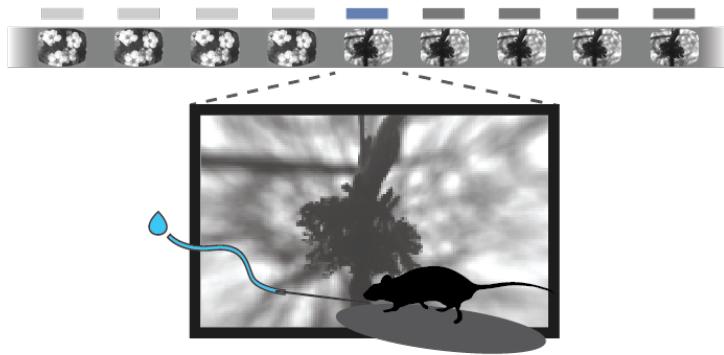


Figure 2.1: visual change-detection task

2.1 Experimental Setup

In the experiment conducted by Allen Brain Institute Neuropixel probes were inserted into the mouse visual cortex (see section 2.3) to record neural activity. While the mice are being presented visual stimuli repeatedly, they either get rewarded for detecting image changes or are exposed to the visual stimuli passively. Each mouse (about 80 in total) performs in two sessions on different days with a clear procedure (see fig. 2.2) after learning how to do the visual guided task. Essentially, a mouse would first be trained on one of two image sets, G or H (see fig. 2.3), then perform on a first recording day in one of the sets and on a second recording day in the respective other set. In most cases (65 of 103) mice are trained on G with first day in G and second in H , but also some cases for combinations $G|HG$ and $H|HG$. The two image sets overlap with two shared images and consist of

¹ data source overview: <http://portal.brain-map.org/explore/circuits/visual-behavior-neuropixels>[23.05.23]

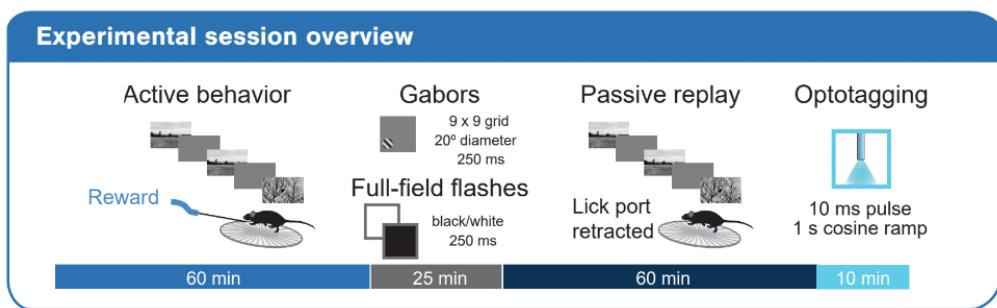


Figure 2.2: single session procedure - 60min change-detection task or active block with natural images as visual stimuli - 25min block of more synthetic stimuli, less relevant for our analysis - passive replay exact same images of active block - 10min block of Optotagging, for identifying cell types see section 2.3

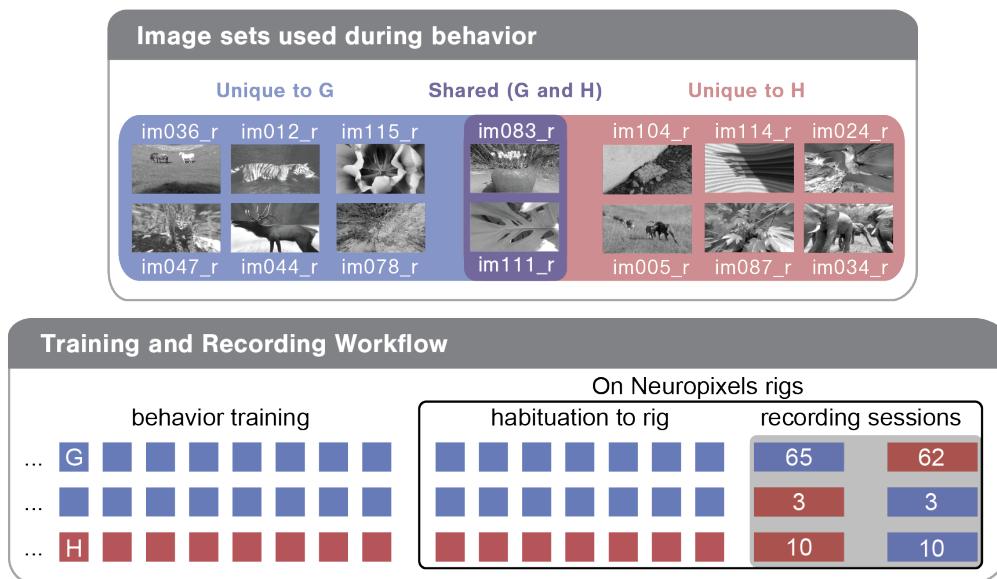


Figure 2.3: imagesets *G* and *H*

8 images each, 14 in total. These natural images are the visual stimuli for the change-detection task where the mouse gets rewarded with water for licking on a spout every time the image changes. For the task Allen Brain Institute constructed a trial structure to acquire whether the mouse successfully detects an image change (hit) or misses it. If it manages to lick on the spout in a short time window after the change, it gets rewarded with water (see fig. 2.2). In detail this is realised by a sequence of stimulus presentations where a single presentation is 250 ms and between each of them are 500 ms of gray screen. After a random number of same stimuli-repeats, the stimuli changes to a different image and a reward becomes possible. Furthermore, there is a 5 percent chance that a stimulus presentation is omitted, resulting in a sequence of same stimuli chains of variable length and variable time between two single presentations.

A session starts with 60 min of task, the active block, continuing with 25 min of simple stimuli, like Gabors and full-field flashes, followed by the passive block; a replay of the exact same sequence

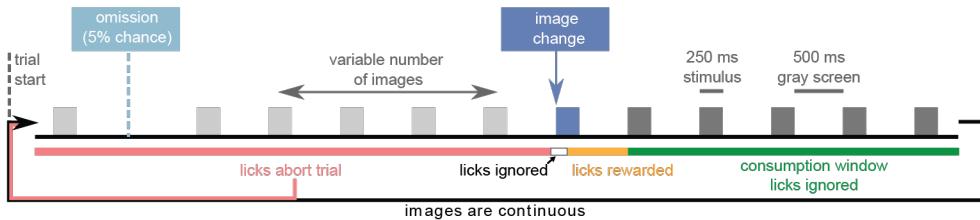


Figure 2.4: inside the task block

of visual stimuli from the first block *without* rewards. Finally, there are 10 min of Optotagging (see section 2.3) to identify certain cells later on.

2.2 Neuropixel Probes

The neuronal response data was obtained with Neuropixel Probes 1.0 (shown in fig. 2.5), which can record electrophysiological signals of various brain regions at the same time. They are developed by imec² and manufactured using CMOS, a process similar to that of chips in phones and cameras, which makes them smaller and less expensive than previous recording devices. A single probe contains 960 recording sites along a narrow shank, allowing the detection of action potentials through 383 data channels at a time from hundreds of neurons simultaneously. Researchers can insert multiple probes into the same brain to observe real-time interactions between different areas of the brain with an unprecedented level of detail. In this case up to six probes record spiking activity in cortex, hippocampus, and thalamus (see next section 2.3 and fig. 2.6 for details).

In opposition to optical physiology methods, individual neurons are only tracked through their action potentials and are referred to as units. Since, essentially, extracellular electrical signals are recorded, it can not be guaranteed that spike times assigned to one unit all originate from a single cell. In that way Neuropixel recordings although faster (and *in vivo*) may not be as accurate for identifying individual cell activity as optical methods (Siegle et al., 2021). Allen Brain Institute is providing the respective optical physiology data, which is not part of our research project but has been used additionally of studies in the past.

2.3 Target Areas, Brain Structures and Cell Types

For the recordings of neural activity six probes have been inserted into the mouse brain, in six distinct cortical brain structures (V1, LM, AL, RL, PM, AM) and two thalamic regions (LGN, LP), as well as hippocampal structures (e.g. CA1). Generally we are focusing our analysis on examining the stability of representations in the primary visual cortex (V1) since it is, as a low-level structure, closest to the sensory input and has been expected to provide stable representations Deitch et al., 2021. Further we are taking higher-order brain areas into account, where signals get integrated. This refers to the hierarchy of brain structures, which has been specified by Siegle et al. (2019) for the structures contained in our data:

² in collaboration with funding partners at the Allen Institute, the Wellcome Trust, the Gatsby Foundation, and HHMI

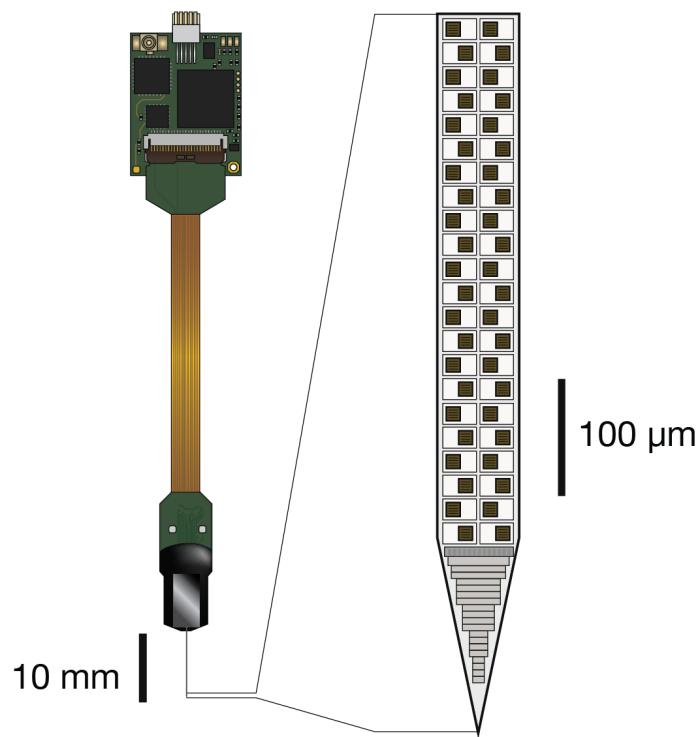


Figure 2.5: Neuropixel 1.0 Probe

Our analysis indicates these areas represent at least seven distinct levels starting with LGN and V1, followed by LM/RL, LP, AL, and finally PM and AM at the highest rungs, [...] (Siegle et al., 2019, p.00)

Through optotagging it is possible to identify specific inhibitory Interneurons in two types of genetically modified mice (Cre+ lines Sst and Vip). In the end of each session a light source is used to trigger action potentials in the cortical target cells which express fluorescent proteins to label them. This gives us the opportunity to compare inhibitory and exhibitory cell types. Considering the relations between brain hierarchies in the visual cortex and cell types to the stability of neural response patterns while performing on a visual-task could help us gain insights into learning processes.

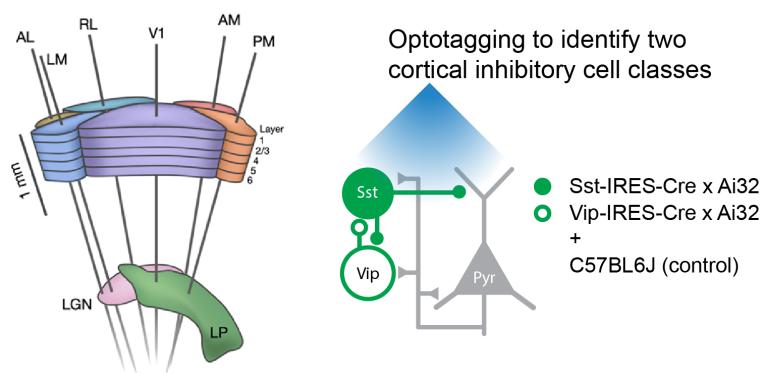


Figure 2.6: Brain Regions and Optotagging

3 Results

In the following sections, we explore findings of representational drift in the mouse visual cortex in detail, breaking it down into three main questions. The first one being if representations systematically change over time (see section 3.1). This has been shown in a number of studies and we conduct a replication in our research project using the recently (2022) published Neuropixel electrophysiology recording of Allen Brain Institute to validate the findings of drift. In this regard we first examine units located in the primary visual cortex (V1) in their activity, comparing between presentations of the same image and then, similar to Deitch et al. (2021), differentiate between hierarchies of brain structures and cell types to identify potential relationships or associations. Subsequently and still reproducing preliminary findings, we examine whether the behavioural state of mice has an impact on the changes in neural activity patterns in response to visual stimuli. Since visual responses, even those in the primary visual cortex, are strongly modulated by the behavioural state Niell and Stryker, 2010. Our approach here is very similar to the work of Sadeh and Clopath (2022) who based their analysis on the previous (2019)¹ large-scale recording of the Allen Brain Observatory viewing experiment. Throughout our work, we define the behavioural state by running speed and pupil size of the animals. Sadeh and Clopath (2022)'s results

[...] suggest that variability of the behavioural state of animal can contribute significantly to changes in representational similarity. (Sadeh and Clopath, 2022)

We therefore ask if representations change systematically with behaviour in the provided dataset (see section 3.2). One major difference of our underlying viewing experiment to the last recording of the Allen Brain Observatory that Sadeh and Clopath (2022) conducted their research on, is the change detection task in the first block of the session, giving us the opportunity to analyse the influence of task-related variables. Finally, our third research question is if representations systematically change with task-related variables (see section 3.3). We are approaching task-dependent factors by analysing first whether the task as its own variable is contributing to representational drift. In the following results we distinguish comparisons of representations within each block as well as across both blocks. This means we explore three cases: how representations change within the active block, within the passive block and from active to passive. When comparing the two blocks, representational similarity is representative of comparing neural activity patterns in response to the same series of visual stimuli in association to a task (active block) versus in absence of the task (passive block). Lastly, we examine task-related data in more detail within the task block. Since the experiment is a relatively constant and well trained task where Neuropixel probes record the same units only throughout a single session, the task-related variables are limited to measurable 'inside-task' variables. We can find challenges and complexity within the same task that manifest in the partly unknown visual stimuli. Therefore, we define the adaption to a new set of images while performing the same change detection task, as a variation of task novelty. Possibly related with the duration of the task, we examined task-relevance by identifying the animals motivation through its

¹ Technical White Paper

engagement levels. Particularly with reward-rates, lick-times and reward-volumes as indicators when approaching the question whether task-relevance contributes to representational drift.

3.1 Representational Drift in the Visual Cortex

Do representations systematically change over time?

The first question we focused our data analysis on is whether we can replicate representational drift in the primary visual cortex with the Allen Brain Institute data sets. Particularly, if representations systematically change over time. Representations refer to the pattern of neural response to a visual stimuli at a specific point in time. We finally identify representational drift through calculating representational similarity (via correlation coefficient, see chapter 5) between neural responses to replays of the same stimuli, since decreases in similarity would indicate changes in representations. The electrophysiology data contains each unit's spiking times and we analyse the ones that pass a quality criteria, proposed by Allen Brain Institute (among others a minimal firing rate, see chapter 5: section 5.1). We define each unit's activity by calculating its firing rate (higher frequency indicates higher activity) within the duration of a stimulus presentation, but instead of looking at single stimulus presentations we conflate sequences of same stimulus presentations until there is a stimulus change to one presentation. The resulting vector of unit activities, also called population response vector, would be a neural representation of a stimulus presentation which we would like to compare to subsequent ones of the same stimulus. Consequently, the population in this case consists of units located in the primary visual cortex (V1 units) which are about 76 units² depending on the animal and session. In the first plot (fig. 3.1) we see the symmetrical representational similarity matrix, a correlation matrix of responses to an image the mouse is familiar with in V1. It is based on an example session, showing us how similar each representation to all other representations of the session is. Accordingly, the stimulus presentations on the x- and y-axis are referencing identical presentations and their respective neural response pattern. On the diagonal the correlation is always 1, the highest possible value indicating perfect similarity since it is the similarity of a representation to itself. If the correlation would be 0, the representations would be uncorrelated and -1 would be a perfect negative correlation. In this case the similarity of two representations varies between 0.4 and 1, suggesting instabilities in neural representations throughout the session. Since the stimulus replays are parted in two blocks (as marked in fig. 3.1), we differentiate between three spaces in our analysis; firstly, within the active block referring to the similarity of representation pairs that are both in response to same stimulus presentations in the first block of the experiment, secondly and likewise, within the passive block with similarity representations of the passive replay block and thirdly, across the two blocks.

For further analysis of representational stability and to return to our question, if representations change systematically over time, we take the time component into account. Therefore, we plot the neural response pattern similarity of the first stimulus presentation to all following replays of that stimulus, by their point in time.

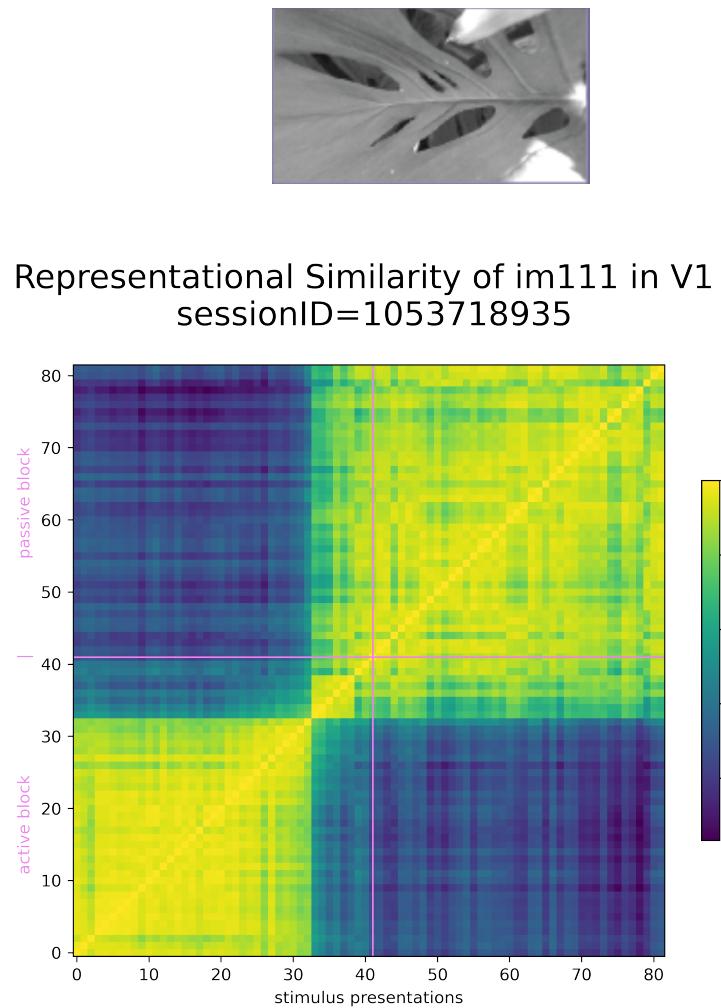


Figure 3.1: On top Image 111 and below the respective correlation coefficient matrix or representational similarity matrix for session 1053718935 in V1, higher values of 1.0 showing highest resemblance between two stimulus response patterns

3.1.1 representations drift across time scales

Following the same example session as the previous figure 3.1, we see how representational similarity develops through one complete session in figure 3.2 with the time in seconds on the x-axis. While the gap in the middle depicts the time between active and passive block, where the stimulus is not being presented, the colours differentiate between the three described spaces: within the active block in blue, green across the blocks and slightly faded (orange) within the passive block. In the top left, the representational similarity starts with 1.0, since it is the similarity of the representation to itself, therefore the first following stimulus replays are within the same block (the active) and the similarity

² varies between 15-121 with 76 units on average

3 Results

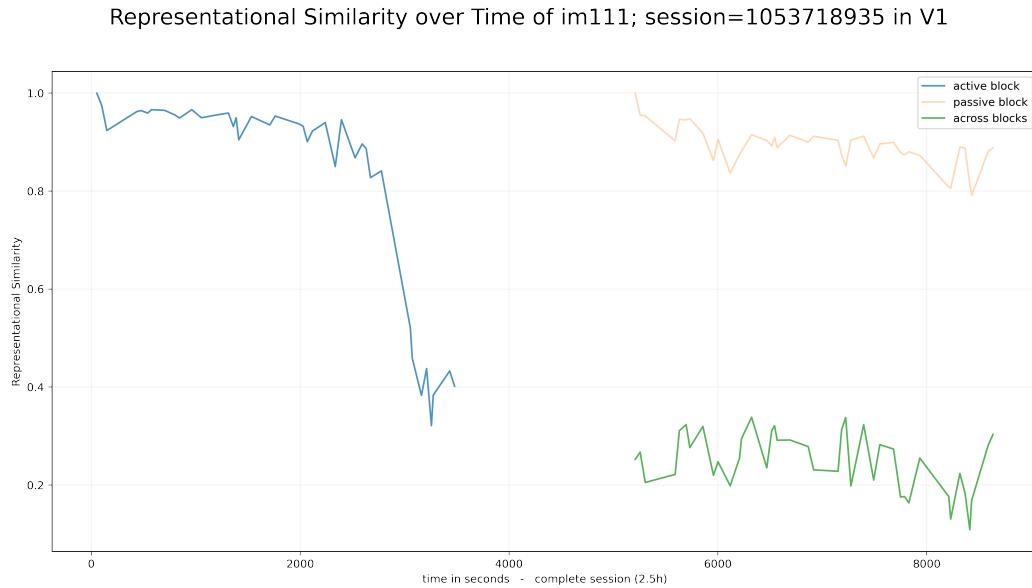


Figure 3.2: Representational similarity over time for example session (1053718935) in V1 and response to image 111

to representations in the second half are across blocks. The curve for representational similarity within the passive block takes the first representation in the second block as reference point instead and will be analysed further in figure 3.3. The graph shows a downward trend over time, which is eminently more conspicuous in the active block than in the passive replay and suggests that representational similarity decreases over time throughout the session at hand. Asking if this trend is being reflected by all sessions in the dataset, we look at fig. 3.3; a similar view, but averaged over all available sessions, marked with the dashed line, while the faded curves insinuate each of the sessions representational similarity progression. The averaged within passive block curve is also shown and coloured in orange instead. Generally the results show decreases in similarity of representations within each and across blocks, significantly more drastic in the block, where change-detections get rewarded and therefore confirming our findings of the example session. Furthermore it might be noteworthy, that representational similarity does not fall below a value of 0.3 in the majority of sessions, although we found single outliers that are close to uncorrelated, slightly negative correlated at times. Respectively the maximal resemblance of neural response patterns we found on the top was about 0.9 with the majority of sessions below 0.8 towards the last stimulus replays of the session. In general representational similarity is decreasing over time, suggesting a systematic change over time. Comparisons of the blocks exhibit higher similarities in the passive blocks, indicating more stable representations over time. Across blocks, which means relatively over a longer timescale, we see the lowest resemblances of representations, while it also shows the slowest decrease of similarity. Measuring neural activity with Neuropixel probes does not allow us to track units over more than a single session or across days, thus we are limited to timescales from seconds to hours. Representational similarity decreases within each block over the time of 60 minutes (fig. 3.2, fig. 3.3) and additionally we can see this decrease throughout the complete session, with up to 2.5 hours between the first population response to a visual stimuli and representation

3 Results

All Sessions: Representational Similarity over Time of im111 in V1 - 96 sessions



Figure 3.3: Representational similarity over time for all sessions (interpolated and averaged) in V1 and response to image 111

of the last replay of the same stimuli, implying that similarity decreases over timescales spanning minutes to hours. Subsequently, the influences of the underlying unit population, will be examined in more detail, varying the composition of units (by location and cell type) to evaluate the similarity of their activity patterns in response to replays of a visual stimuli.

3.1.2 representations drift across hierarchy

The primary visual cortex, or V1, as a low-level brain structure processes sensory information from the eyes and was originally considered to represent visual stimuli more stably than higher-level structures, such as CA1 in hippocampus. However recent work (e.g. Deitch et al., 2021) found representational drift in multiple visual areas and that neural-code stability did not reflect the hierarchy of information flow across areas. Our data seems to validate these findings, as we pointed out before representational drift occurs in V1 and subsequent we are comparing the drift following the information flow to more integrative, higher-level brain areas of the hierarchy. In fig. 3.4, which is based on the same example session, we see a similarity of representations distinguished by colours for three brain structure hierarchies. Neural response pattern similarity decreases for units located in primary visual cortex (V1), in the remaining visual cortex ($VIS \setminus V1$)³, as well as in the hippocampus (CA1), indicating that representations drift across hierarchy levels. Particularly within the active and across blocks CA1 units representations constantly seem to show higher resemblances than the upstream brain area $VIS \setminus V1$, while sharing a very similar decrease. V1 units' similarity on the other hand fluctuates a lot more, which may be due to being a significantly smaller population of 57 units (287 in $VIS \setminus V1$ and 197 in CA1). They appear initially to represent

³ The primary visual cortex (V1) being a sub structure of the visual cortex (VIS), can be seen as sets of units and thus the expression refers to VIS excluding V1

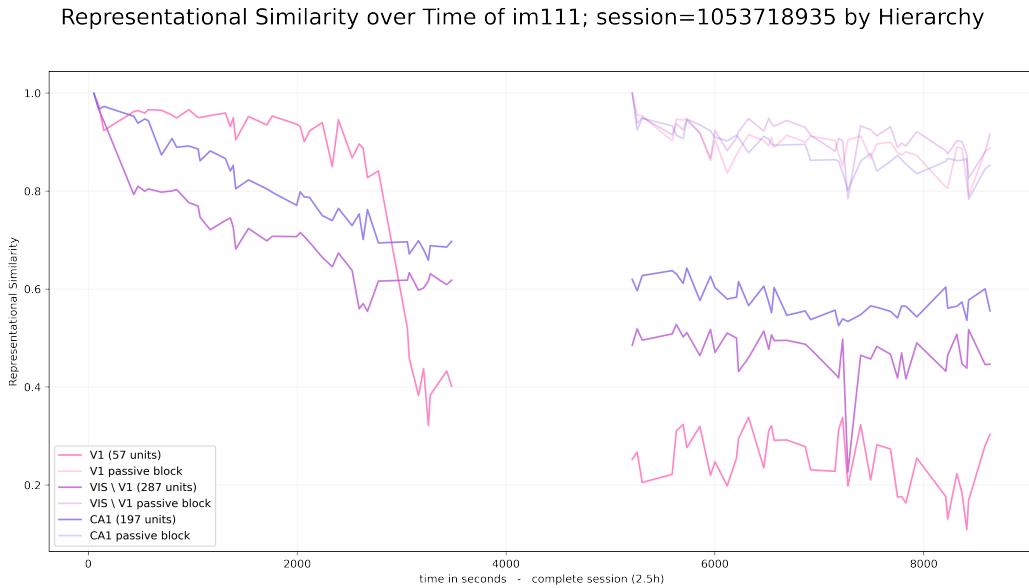


Figure 3.4: Single session's (1053718935) representational similarity by hierarchy (V1, VIS excl. V1 and CA1) in response to image 111

visual stimuli more stably however as time progresses there emerges a lower resemblances between representations towards the end of the active block and across blocks over time. Ultimately we see a sudden drop within the active block and significantly lower representational similarities than the two higher-level brain structures. However within the passive block the different hierarchies stay closer together and the representations in visual cortex excluding V1 appear slightly more stable. Additionally we can observe that the curves, distinct in hierarchy, show overall drops, peaks and dips in a similar manner. Looking at all available sessions, figure 3.5 suggests, that neural response patterns are at all times and regardless of block space more stable in CA1 units than in VIS \ V1 units, while V1 still varies from showing a slower decrease in the beginning of the session, then dropping lowest until approximating the representational similarity of the remaining VIS units. However this is clearest within the passive block opposed to our example session and about the first 20 minutes (within active block), where the graphs are not as distinguishable. Consequently, we can find drift in all brain areas, irrespective of hierarchy, while the hippocampus seems more stable than upstream areas and the primary visual cortex less stably decreasing.

3.1.3 representations drift across cell types

The sharpest distinctions of stability in neural response patterns, we found between different cell types in the visual cortex. Analogical to the previous section we varied the underlying unit population to compare decreases in representational similarity over time dependent on the cell type instead of the brain structure. The optotagging (as described in section 2.3) gives the opportunity to identify and tag inhibitory cells, even though electrophysiology data merely consists of electrical signals. The remaining untagged cells and vast majority will be considered as exhibitory cells. In figure 3.6 we see how similar representations are over time (in our example session) after

3 Results

All Sessions: Representational Similarity over Time by Hierarchy - 96 sessions

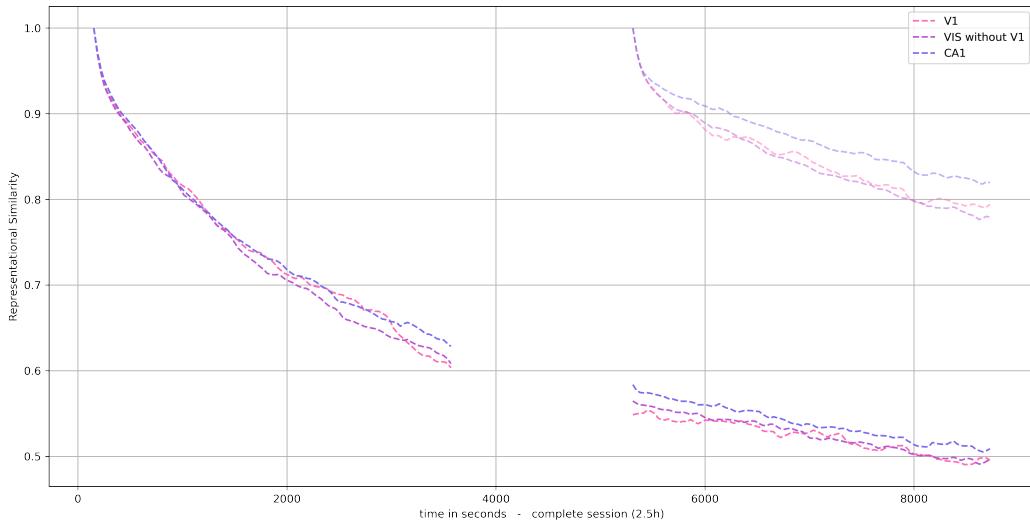


Figure 3.5: All sessions representational similarity by hierarchy (V1, VIS exclud. V1 and CA1) in response to image 111

Representational Similarity over Time of im111; session=1053718935 in VIS cells inhibitory vs. exhibitory

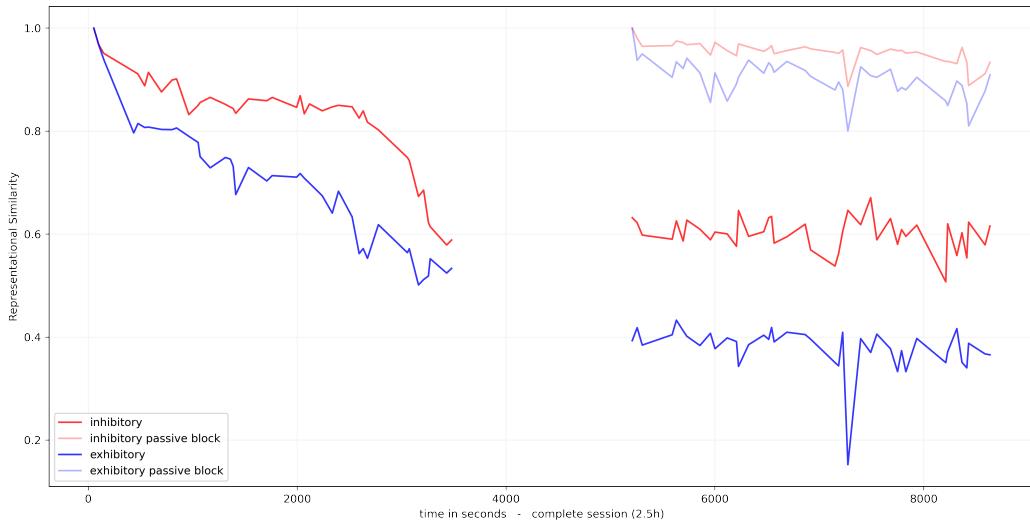


Figure 3.6: Single session's (1053718935) representational similarity by cell types (inhibitory vs. exhibitory) in VIS

dividing the unit population in inhibitory tagged cells (coloured red) and exhibitory cells (coloured blue), insinuated again slightly more faded are representations within the passive replay block. Markedly the two cell types differ in their level of representational stability, response patterns in exhibitory cells showing less resemblance than inhibitory responses. While representational

similarity of exhibitory cells decreases within the active block by almost 0.5, it drops further across blocks, gradually shrinking, though fluctuating, below a similarity value of 0.4. The tagged cells in contrast, show a slower decreasing resemblance of response patterns at first (within the active block), intensifying in the last 10-15 minutes of the first block and alternate around a representational similarity value of 0.6 across the blocks. Within the passive block however the divergence between the cell types is more subtle, still depicting a steady decline of representational similarity. Partly

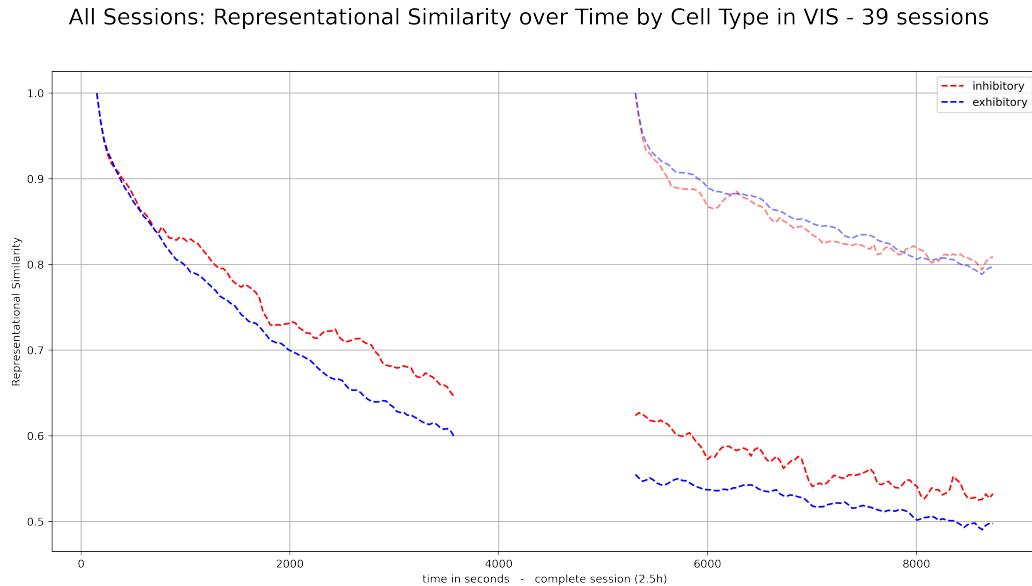


Figure 3.7: All sessions representational similarity by cell types (inhibitory vs. exhibitory) in VIS

validating associations of the example session plot (fig. 3.6), we see in figure 3.7 the same plot generalised over all available sessions, which are only 39 in this case, according to the number of sessions with genetically modified mice and where we were actually able to tag inhibitory cells through optotagging. Similarity of representations are decreasing faster in the inhibitory cells within the passive block and the example session appears to be an exception in that regard, since the curves decline conversely (fig. 3.6). Additionally it is noteworthy, that this is also opposed to the order within the active and across blocks. Visibly the averaged representational similarity curve of the exhibitory cells is smoother, which might be attributable to being the result of a larger amount of units (fig. 3.7). Our results suggest, that representations of visual stimuli change over time across (both) cell types. In the visual cortex units tagged as inhibitory cells seem to be generally more stably representing visual stimuli than the untagged, mostly exhibitory cells. Notably the results indicate, that this behaviour is in fact reversed if we compare representations solely within the passive block.

3.2 Representational Drift Depends on the Behavioural State

Do representations systematically change with behaviour?

The findings above suggest that the neural representation of stimuli is changing over time, it is crucial however to acknowledge the potential influence of other variables, including the behavioural state of the mouse. Since changes in behaviour can modulate neural activity (Niell and Stryker, 2010), a contribution to representational drift is worth considering. Thus recent work showed that changes in neural response patterns should not solely be characterised in relation to stimuli, as described by Sadeh and Clopath (2022). Their results suggest, that parameters like the behavioural state of the animal can modulate the representational similarity across repeats of the same stimulus. With our analysis we can confirm these findings in our extended dataset. Based on the work presented by Sadeh and Clopath (2022), we analyse the behavioural variables running speed and pupil area z-scored (see Methods chapter 5), to examine resemblance of representations in relation to behaviour. We z-scored in order to quantify running speed and pupil area as a single behaviour variable in form of a behaviour vector. This is illustrated in figure 3.8, where we see a plot consistent with

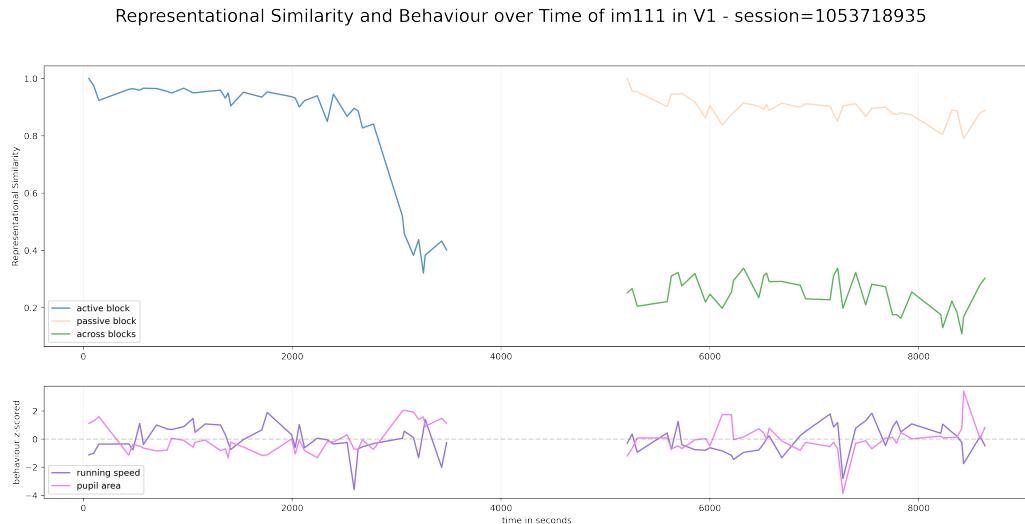


Figure 3.8: single session's (1053718935) representational similarity aligned with behaviour variables (z-scored: running speed and pupil area) below

the example session of the previous section, contrasting representational similarity throughout the session with the z-scored behaviour (in the bottom subplot) over time. While the upper plot was described in section 3.1.1, the bottom plot extends the view, as we have seen it in fig. 3.2, aligned with the z-scored behaviour, each component delimited by a different colour (running speed in purple and the pupil area in pink). In order to get a more coherent visual representation of the relationship between representational similarity and the behavioural state of the animal, we plot all representation pairs similarity dependent on the corresponding difference in behavioural states (ΔB , see Methods section 5.2). This is embodied by the main plot in figure 3.9, while the other two subplots illustrate the marginal distributions of each variable (ΔB and RS). The difference (ΔB) of two behavioural states is calculated by the distance of the particular Z-vectors and the similarity can be extracted from the correlation matrix (see Methods section 5.2). In the scatter

3 Results

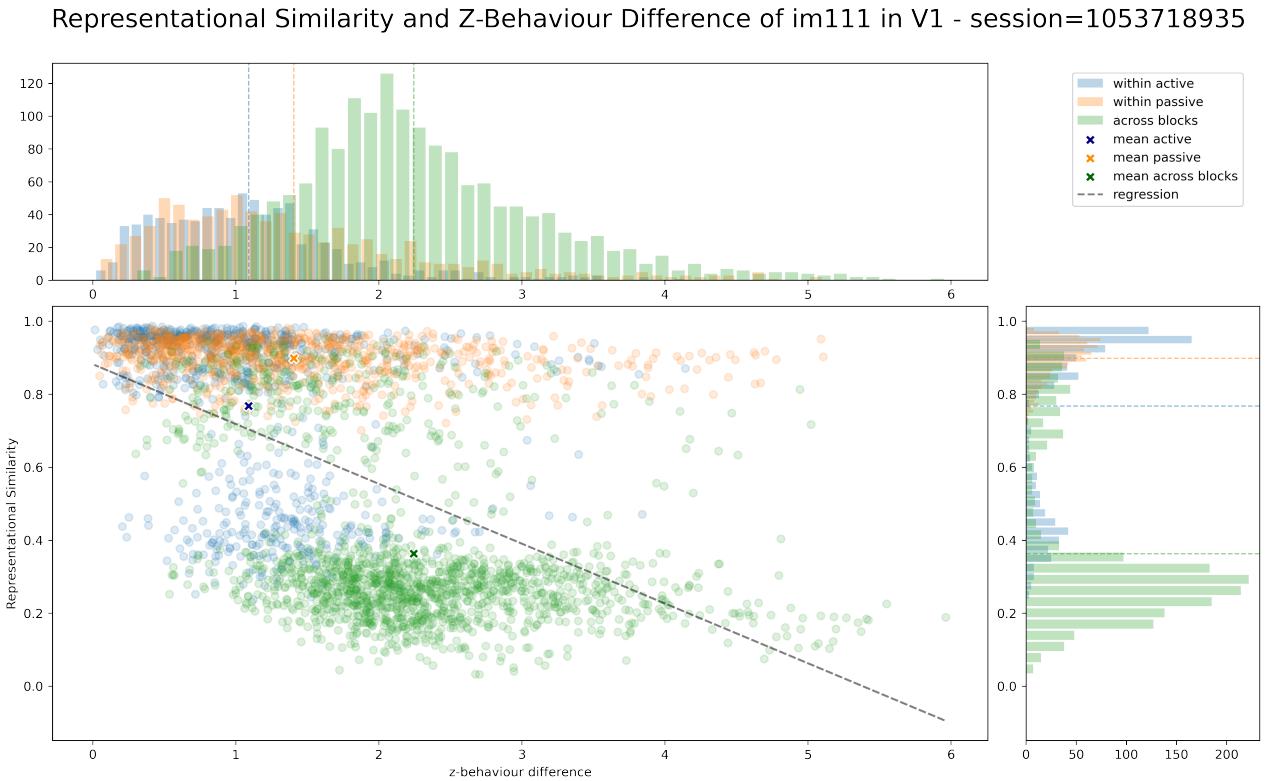


Figure 3.9: single session's (1053718935) representational similarity in relation to behaviour difference with marginal distributions between representation pairs throughout session in V1 and response to image 111

plot of figure 3.9, where we still differentiate between the spaces active, passive and across blocks, we further visualise the overall trend of the data by drawing in the regression line as well as each spaces mean (marked with a cross). While the within active and passive spaces share the same number of datapoints, there is noticeably more than twice as many across the blocks. The regression line decreases visibly, mainly as a consequence of representation pairs across blocks, indicating a negative correlation between representational stability and behaviour variables running speed and pupil size. To enhance generalisation we extend this view by all available sessions of our dataset. Consequently, the scatter plot in figure 3.10 is the same as the previous (fig. 3.9), including data of 95 additional sessions. For visibility reasons we only plot every sixth data point, while the regression lines are calculated on the total amount of representation pairs and the marginal plots still show the entire distribution. The results suggest that with increased differences in pupil size and running speed, the similarity of neural representations in V1 gradually decline. Validating our single-session findings as well as Sadeh and Clopath (2022)'s work, we found a decrease in representational similarity with increased behavioural discrepancies, implying that changes in the behavioural state are contributing to representational drift.

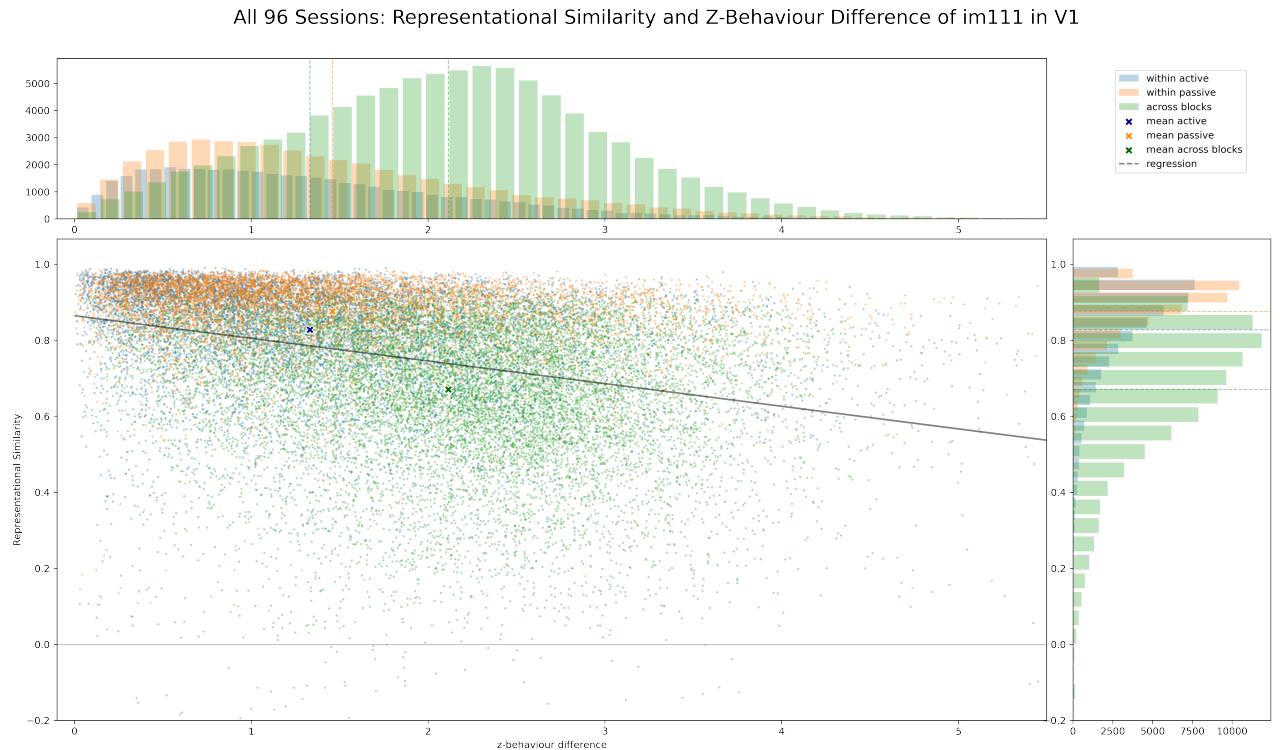


Figure 3.10: all session representational similarity in relation to behaviour difference with marginal distributions between representation pairs throughout session in V1 and response to image 111, with every 6-th datapoint plotted for visibility

3.3 Representational Drift is Impacted by Task-Related Variables

Do representations systematically change with task-related variables?

Taking a step further we want to explore whether task-related variables contribute to drift in representations. Similar to behaviour variables, task-related variables, such as attention and motivation can modulate neural activity (Tang et al., 2023) and since it appears to be a significant influence on representational stability it might be applicable. Before we get into detail with specific task-related variables, we revisit noted distinctions between the active and the passive block, whereas active refers to the (trained) change-detection task, including water rewards as a characteristic opposed to the passive replay block.

3.3.1 change detection task environment influences representational drift

As we have seen in the previous sections we find differences in stability of representations throughout the active versus the passive blocks of the experiment. Remarkably representations within the active block show lower resemblances to each other and therefore appear to drift faster (fig. 3.2, fig. 3.3). In figure 3.11 we contrast both blocks representational similarities directly, omitting across block representations. Between brain hierarchies (V1, VIS \ V1 and CA1) within the passive block of the example session we see primarily, that the curves exhibit a higher degree of proximity in their

3 Results

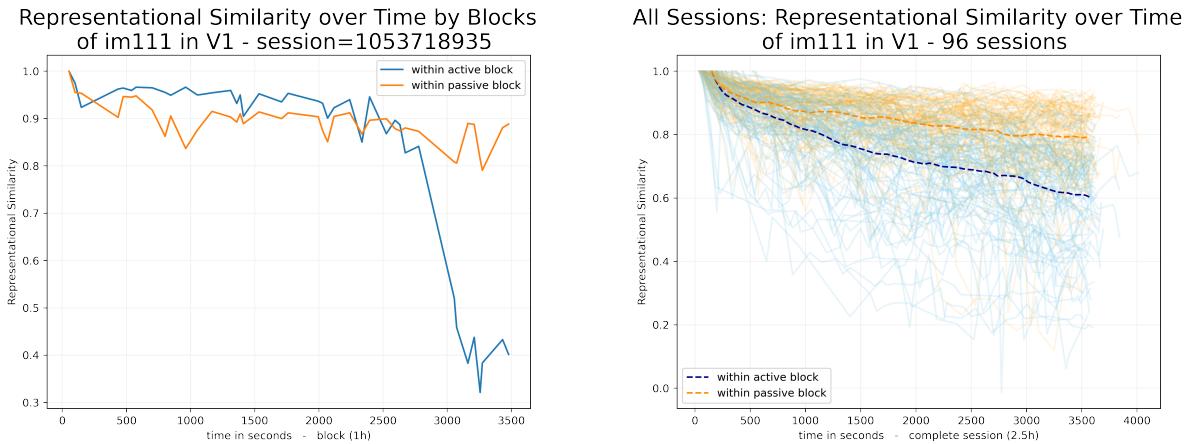


Figure 3.11: Single session (1053718935) plot left and all sessions plot on the right contrasting representational similarity in active to active and passive to passive, both cases in V1 and response to image 111

decrease (fig. 3.4). The trend of reduced separation between them does not persist as additional sessions are considered, in fact a higher stability in the hippocampus (top-level in hierarchy) within the passive block compared to the active block emerges (fig. 3.5). Looking at representations of inhibitory and excitatory cells, we find similar relations between drifts in the example session, as in a clearly reduced separation between each cell types decrease in representational similarity within the passive block (fig. 3.6). Again taking all sessions into account, the observed trend endures. It is remarkable however, that representations in inhibitory cells seem less stable over time within the passive block, while showing significantly higher stability in representations within the active and across blocks compared to excitatory response patterns (fig. 3.7). Even in relation to behaviour our results indicate a higher, though still modest, correlation within the active block than the passive block. This is more visible in figure (3.12), where we plot representational similarity dependent

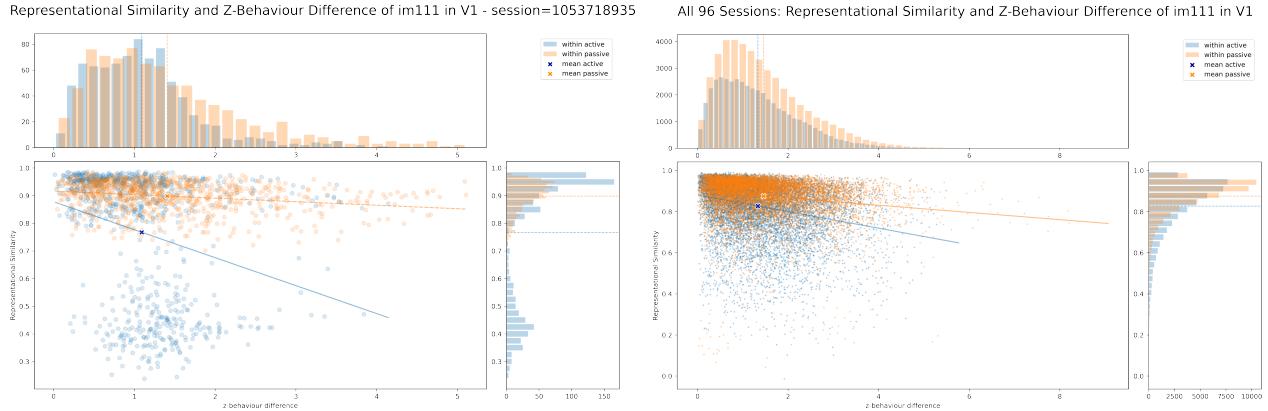


Figure 3.12: Single session (1053718935) plot left and all sessions plot on the right contrasting representational similarity in relation to behaviour for active to active and passive to passive, both cases in V1 and response to image 111

on difference in the behavioural state, without the across blocks response patterns (which is a modification of the plots in the previous section 4.2, fig. 3.9, fig. 3.10). On the left we see the example

session and on the right side all available sessions, where it seems like the mean behaviour difference between representation for distinct points in time for the respective blocks are closely situated. Despite their close proximity, the two means correlation values diverge significantly, which could indicate, that representations within the passive block (also in absence of task) are less impacted by the behavioural state (than within the task-block).

3.3.2 absence of significant influence of stimulus novelty on representational drift

In the results so far we only looked at familiar visual stimuli, images the mice have trained the change-detection task on. Does the representational stability of unknown visual stimuli differ from the familiar ones? Trying to answer this question we take alternative visual stimuli from the experiment into consideration. In this regard we compare stability of neural activity patterns in response to different image presentations. We look at how this is expressed in our example session, which only includes familiar visual stimuli, with a novel session, sharing two familiar images ('im111' and 'im083', see image sets in fig. 2.3). In the familiar session (fig. 3.13), known

Representational Similarity of all Images in V1 by familiarity - familiar session=1053718935

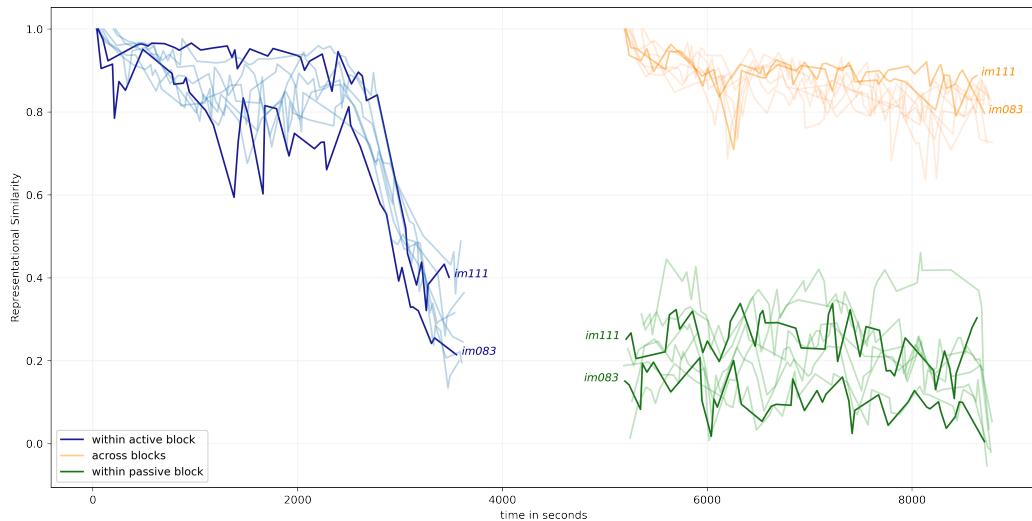


Figure 3.13: Familiar session's (1053718935) representational similarity in response to each image in V1

from the results above, we can see representational similarity varies for different images without notable or especially unexpected anomalies. Within the active block we find almost all images with representational similarity values in between *im111* and *im083* (fig. 2.3), which are slightly higher towards the end of the session, when calculating resemblance within the passive block. For representations across blocks however, we note more variations and alternations in general. Thus figure 3.13 serves as an appropriate (example) reference plot for evaluating a session with stimuli-novelty. Figure 3.13 depicts generally higher representational similarity values over time. Furthermore it does not appear to make a difference whether an image is novel or familiar, as the similarity curves of familiar stimuli do not appear to decrease more or less compared to the

3 Results

Representational Similarity of all Images in V1 by familiarity - novel session=1124285719

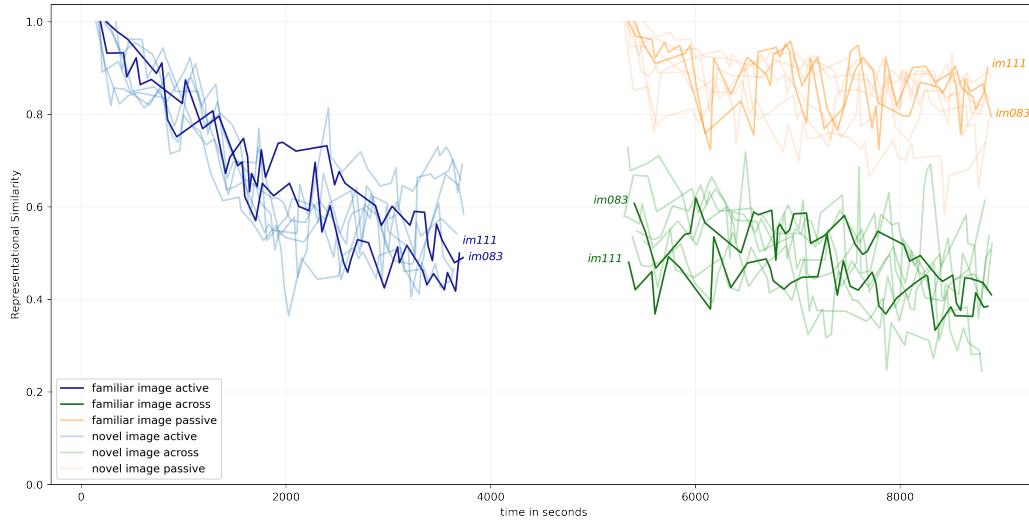


Figure 3.14: Novel session's (1124285719) representational similarity in response to each image in V1

novel stimuli representation similarity (over time). We test this association with all available 41 novel sessions, through picking one familiar (*im111*) and one novel image (*im114*) to calculate each corresponding mean representational similarity curve. The resulting plot (fig. 3.15) is analogous to

All Sessions: Representational Similarity over Time by Novelty in V1 - 40 sessions

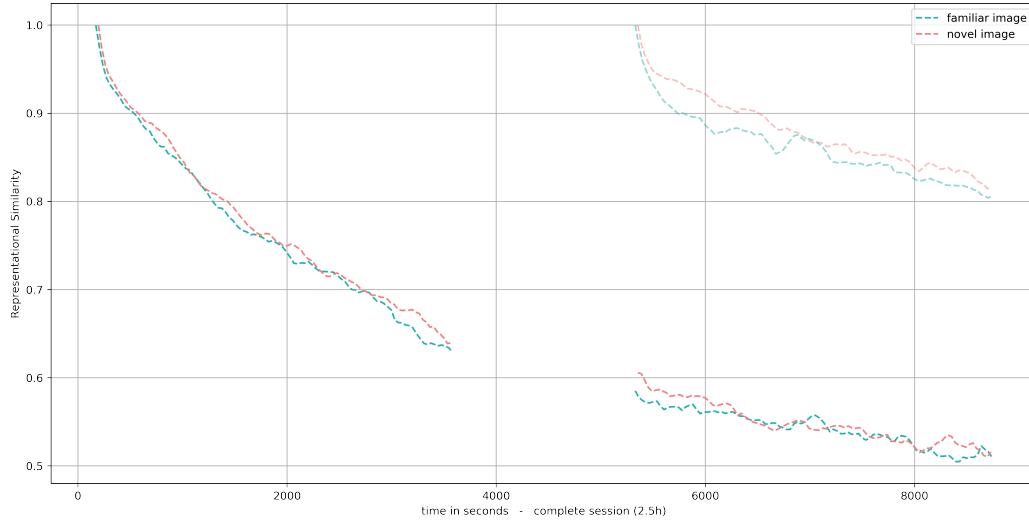


Figure 3.15: all sessions representational similarity by stimuli novelty in V1, comparing two images (*im111* as familiar and *im114* as novel)

the figures in section 3.1 and suggests that stimulus familiarity does not yield reliably increased

stability of representations over time. Contrary to expectations, novel stimuli (in this case) even exhibit a trend of greater representational similarity.

3.3.3 potential influence of task relevance on representational drift

Starting from the premise that task relevance is driven by motivation, which is further influenced by rewards within the task, we can analyse task-related variables such as reward timings and volumes into our analysis. This will allow us to investigate the impact of task relevance on representational drift. Since the mice enter the experiment in a 'thirsty' state, we can assume, in accordance with the premise, that their motivation to receive water rewards and perform tasks is at its maximum at the beginning of the session. Our final figure focuses therefore on reward and performance related variables aligning them based on the example session's representational similarity in the primary visual cortex. Figure 3.16 gives an overview over the course of a change-detection task

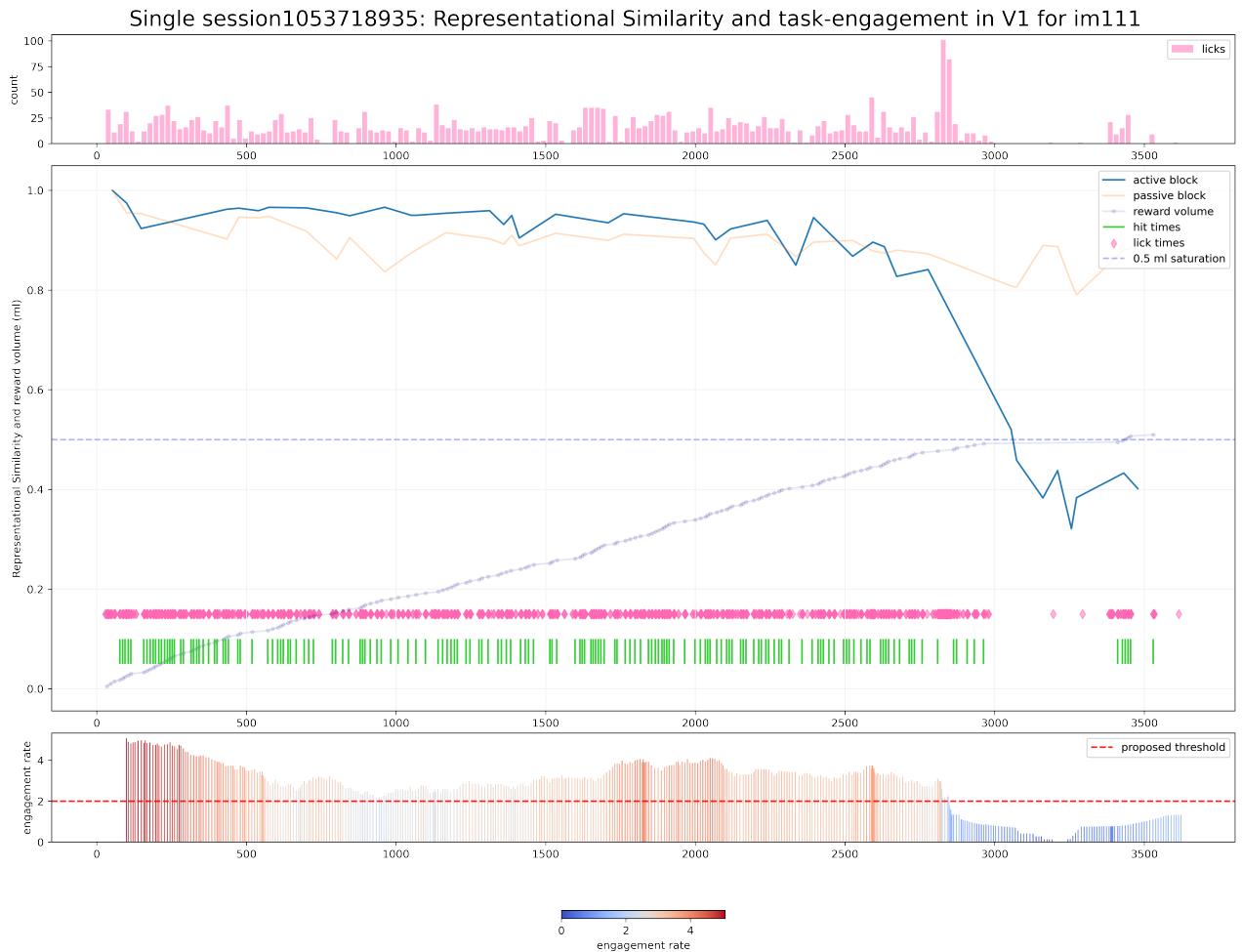


Figure 3.16: Single session's (1053718935) representation similarity over time with performance, reward volumes, engagement, including lick times

block, containing various task-related information mapped on one time axis with representational similarity in the main plot (blue curve), while the within passive curve (orange) is just a projection

3 Results

of the passive replay that follows this first part of the session. Still in the main plot we see pink markers for lick times and green markers for when the lick time was actually the correct response to an image change and therefore a 'hit'. Additionally the lick times are aggregated in the upper plot, indicating how often the mouse licks on the spout over time. Also aggregated we see the amount of water the mouse drinks/ gets rewarded with (in ml) throughout the experiment and the saturation threshold of 0.5ml corresponds to this volume, suggesting a potential saturation with water. In the lower plot we depict the engagement rate that is determined by the task performance (see Methods: section 5.2). Visibly the mouse shows lower engagement towards the end of the task, reflected by the performance indicators as well as lick times, accordingly the task-relevance might decrease over time, exhibited further by the aggregated reward volume approximating the saturation threshold in this particular session.

4 Discussion

In this chapter we are discussing the previous findings, in order to conclude the relevance for understanding the mechanisms of representational drift in mouse visual cortex. Consistent with the results chapter (3), we go through the same categories, reproducing drift in our dataset based on Deitch et al. (2021), as well as reproducing a correlation with the behavioural state as proposed by Sadeh and Clopath (2022) and finally extending the analysis by exploring task-related variables. More accurately we identified task-novelty, task-duration and task-relevance as observable (and examinable) variables, based on the experiment and available data.

4.1 Gradual Changes in Neuronal Response Patterns

Our analysis showed representations changing over time, indicating that they do systematically and across timescales as well as brain hierarchies and cell types. By finding decreases in similarity of representations over all these cases, we can generally confirm recent research on occurrence of representational drift in mouse visual cortex (as well as hippocampus) in our data. In contrast to the subsequent sections, neural representations over time were solely characterised in the stimulus space, varying the underlying unit population exposed to replays of the same visual stimulus over time.

time scales

We observed representational drift over times scales spanning minutes to hours, congruent with the findings of current research. However, our analysis was constrained to individual sessions, imposing a time limit typically no longer than a few hours. With use of the second (optical) dataset tracking specific neurons over longer time scales would be possible (Sheintuch et al., 2017) and might be used to even out disadvantages of this technology. This was done by Deitch et al. (2021), who only found small differences between their analyses of the electrophysiology and optical data of the same experiment.

hierarchy

If we ask how the brain hierarchy can influence changes in neural response patterns, recent work found, contrary to the initial expectation, that lower level structures are not necessarily more stable than higher-level structures.

If anything, our analysis shows that the coding stability of some cortical (V1 and LM) and subcortical (LGN and LP) areas exhibit an opposite trend with respect to their hierarchy.
(Deitch et al., 2021, p.4334)

Our comparison of representational similarity decreases between hierarchies, highlights this notion, since we found the highest response stability in CA1 units, which is as a structure located in the

hippocampus (crucial for memory formation) and therefore an upstream brain area from the visual cortex (fig. 3.5). The primary visual cortex as an early stage of sensory processing and lower-level than the rest of the visual cortex, however shows higher variability in its representational drift over time, decreasing slower than in VIS\CA1 at first and then faster again, making it difficult to arrive at a concise conclusion. Especially since the higher variability could also be a consequence of V1 embodying the smallest analysed unit population. Given this perspective and in anticipation of section 4.3, we could derive the question, whether V1 exhibits a greater dependence on task-related variables compared to the remaining regions of VIS and CA1.

cell types

In the case of representational stability induced by cell types, we found the most prominent distinction. Although exhibiting drift as expected, inhibitory cells depicted clearly more stability than excitatory cells throughout the session. Surprisingly however is, that this is reversed within the passive replay of the experiment, but not across the complete session, which could indicate a shift, requiring further investigation.

4.2 Impact of Behavioural Variables

In order to better understand the mechanisms of representational drift, we aimed to investigate the relationship between representational similarity and two behavioural variables, namely running speed and pupil area, over time. Our findings revealed a significant correlation between representational similarity and differences in these behavioural variables. These results align with previous research of Sadeh and Clopath (2022), who initiated a broader exploration contributing factors beyond the visual stimuli, like behavioural variables contributing to drift. Although they emphasise:

[...] we observed large changes in representational similarity between blocks of stimulus presentation with strong behavioural changes. But this effect can be confounded by the passage of time between the blocks, which may lead to other sources of variability such as changes in the excitability of neurons for instance. (Sadeh and Clopath, 2022, p.5)

Overall, our study seems to verify this relation by demonstrating a correlation between representational similarity and behavioural variables. While consistent with prior literature, it is important to recognise that a significant portion of this relationship may be influenced by factors related to across-block differences rather than solely the behavioural changes themselves.

4.3 Impact of the Change-Detection Task

In the results concerning task-related variables we started coarse, finding that the task generally influenced neural activity patterns in mouse visual cortex differently than in task-absence. Then proceeded to zoom into the (active) change-detection task block to examine more concrete task-related variables - task-novelty and task-relevance. While task-novelty was deducible more directly from our data task-relevance involves a bit more interpretation. Implicitly we also examined task-duration, since representational similarity decreases considerably over time throughout the task. However the results at hand only provide limited foundation for a more advanced understanding of

the individual contributions of task-related variables to the decrease in representational similarity. We are going to argue though that task-duration effects on representational similarity is likely induced by task-relevance effects in our experiment, considering that possible cues for both could be attention, motivation and engagement levels.

task or no task

Different stabilities in the task-block (active) and the no-task block (passive) suggests that the task itself influences how fast representations change. The visible difference here indicates how indeed task-related variables could contribute to changes in neural response patterns. Recent research suggests, that the most prominent differences in representational similarity are across the blocks and also found in the optical dataset (Deitch et al., 2021, as cited in Sadeh and Clopath, 2022[p.4]). Our results can confirm this observations, although we find the highest decrease within the active block. Faster decrease within the task block opposed to the replay block, could indicate that there is a correlation between task-related variables and representational drift additional to passage of time and influences of the behavioural state, since both blocks describe the same time span, as well as the exact same sequence of stimulus presentation. We also found similar behavioural differences within each block, opposed to across blocks. Nevertheless this relationship was not thoroughly examined in our analysis and would need further investigation in future studies.

task novelty and stimulus variation

We compared representational similarity in response to trained familiar visual stimuli with the ones from image sets that the mice have never seen before and therefore represent task novelty. This means it is not the task that changed but the images mice decide on whether there has been a change and which the reward depends on. The results suggest that novel stimuli have no particular impact on representational drift, as the decrease of representational similarity does not appear to be higher or lower in relation to the familiar stimuli. However novel images seem to be represented with more fluctuation in their similarity to each other. From performing a 'skill transfer' to unknown visual stimuli in a generally constant task, we would expect a higher cognitive load, but our results do not seem to be constituting this notion. In our analysis we also only compare one novel-familiar image pair over all novel session instead of examining the relation between all visual stimuli pairs, since the single sessions we looked at didn't indicate a clear pattern. Consequently the analysis of this relation over all sessions is not yet concluded in that regard. Furthermore it might be relevant to take extent the analysis by taking image similarities into account.

task-relevance: attention, engagement and motivation

Approaching impacts of task-duration and/or task-relevance, we found that mice engaged decreasingly over the course of the task reflecting the decline in representational similarity. It seems that the main drops in similarity of representations occurred after the animals reached a reward volume of 0.05ml, which can be interpreted as the mice are losing motivation when they get satiated with water. This would indicate the task is losing its relevance from a thirsty mouse in the beginning to a satiated mouse towards the end. Supported by our evaluations of engagement serving as our indicator for motivation, possibly for attention as well. In that manner we defined engagement by task performance, as proposed by Allen Brain Institute and additionally by lick times unconstrained

to success. The Results however, show this relations qualitatively for a single session, which makes a general conclusion impossible, although still worth discussing. Following our interpretation lower motivation due to being less thirsty can make representations associated with the task become less stable and drift more over time. Being thirsty could particularly underline the aspect of task-relevance, while motivation could also decrease as consequence of task-duration. Consequently our results would provide indications that large decreases in representational similarity over time could depend primarily on motivation and attention, which we can only speculate at this point, since we did not quantify these possible correlations further. Similarly the available eye tracking data might give more information about the attention of the animals throughout a session, but would require further analysis. Ultimately and similar to the conclusion of Sadeh and Clopath, 2022, a contribution of task-related variables should not be disregarded when trying to shed light on the mechanisms of representational drift and neural code.

4.3.1 Prospect

Overall we reproduced representational drift in our electrophysiology data for the visual cortex, as well as its correlation to behavioural variables and identified that task-related variables play a role in contribution to drift. It would be interesting however to train a classifier to examine when (or how long) images are distinguishable, based on the neural activity within different populations of units. We also did not include (due to time constraints inherent in a bachelor's project) the tuning of neurons in our analysis, which could provide additional insights.

5 Methods

Universally the method of this research project is a data analysis, crucial and commonly used for quantifying, comparing, and interpreting empirical data in the study of neural code. In order to examine different variables contribution to representational drift, the core of this analysis is processing neuronal data to quantify representations. This is the data preprocessing, described in the next section (section 5.1) and needed to enable the analysis of their stability under influences of various factors (see section 5.2). While the partially processed electrophysiology data was accessed via *AllenSDK* and downloaded with default functions, dataprocessing and analysis was carried out with custom-written Python functions in *JupyterNotebook* and can be found on *Github* (<https://github.com/9-cordes/Bachelor-Project>).

Most variables and parameters can be extracted directly from the SDK. Additionally Allen Brain Institute provides a repository¹ with *Notebooks*, including explanations and first-level functions and covers for instance, how to identify inhibitory cells with optotagging. In the following sections we will therefore focus on the more complex parts of our data analysis.

5.1 Data Preprocessing

In this section we will deconstruct the neural representations, as they are the foundation for our data analysis. Representations are more specifically the neural response patterns to natural images. The response refers to the unit activity during presentation of the specific visual stimulus. For a population of units we therefore compute a unit activity vector as the response pattern, also called population response vector. Before selecting any unit population (e.g. by brain structure or cell type) we filter units by quality criteria, recommended by Allen Brain Institute:

<i>quality == 'good'</i>	(quality can either be 'good' or 'noise')
<i>firing rate > 1 Hz</i>	(mean firing rate over entire recording)
<i>snr > 1</i>	(signal-to-noise ratio for 1D waveform)
<i>isi violations < 1</i>	(ratio of refractory violation rate to total spike rate)

Constructing a Population Response Matrix

In order to retrieve each units activity we extract the number of spikes during a single stimulus presentation and divide it by the duration (approximately 250 ms) to get the firing rate, this process is similar to binning. Carrying it out for each stimulus presentation, we receive a matrix of unit activity vectors, each the respective response to every visual stimulus of the session. As illustrated in figure 5.1 this also includes the omitted pictures, marked by the colour blocks below the unit activities.

¹ GitHub - <http://portal.brain-map.org/explore/circuits/visual-behavior-neuropixels>[23.05.23]

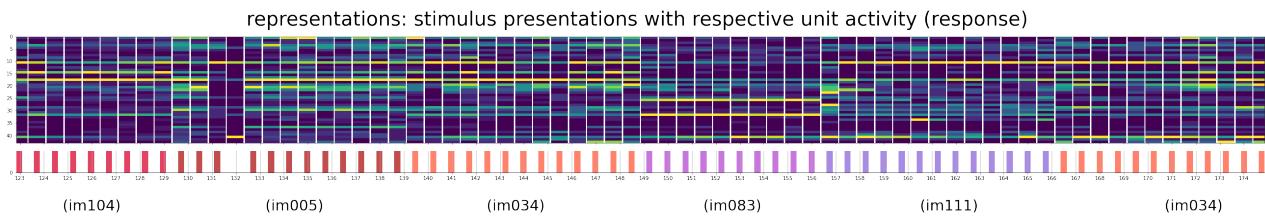


Figure 5.1: 70 unit activity vectors/ responses to visual stimuli, where yellow indicates high activity and blue low (see also fig. 2.4)

As last preprocessing step, we reduced the amount of unit activity vectors by calculating chain averages. We defined a chain as a same-image sub-sequence of stimulus presentations. This means as long as there is no image change we are describing the same chain and take the mean firing rates as the unit activity vector for the respective visual stimulus. This is supposed to deflate the neural response data, so that every (chain averaged) representation is now across the length of multiple original (250 ms) presentations, already sharing a very similar activity pattern. Consequently we get for a unit population of choice the respective activities aggregated over identical stimuli and when we filter this for a specific visual stimuli (for instance image 111) we get all response patterns for replays of this stimuli. In essence these are the neural representations of the corresponding stimulus for a subset of units throughout the session and therefore fundamental for analysing the stability in relation to various factors.

5.2 Data Analysis

The core data analysis consists coarsely of three parts, the first is about quantifying representational drift, the second about associations with behavioural variables and the third takes task-related variables into account. Following this structure, we are providing more detailed descriptions for those processing methods that can not be derived from openly accessible *Allen-Workbooks*. Examining whether representations change over time, we measured how similar neural response patterns to repeats of the same visual stimuli are to each other.

Representational Similarity

To quantify representational drift we computed similarity between representations and therefore took Pearson's correlation coefficient, which results in a value between -1 and 1 as a similarity measurement.

Pearson correlation coefficient:

$$\rho_{ij} = \frac{\text{cov}(v_i, v_j)}{\sigma_{v_i} \sigma_{v_j}} \quad (5.1)$$

If we calculate the correlation coefficient between all pairs of representations, this results into a correlation matrix, where we can simply read out the population response similarity of two stimulus presentations. Accordingly a representation can be mapped back on the time axis discretely, by calculating the median of the stimulus presentation duration. Similarly we map it on behavioural..,?

behavioural variables

z-scoring behaviour, where σ is the standard deviation, t is a point in time, r_t is the running speed at t and $\langle r_t \rangle$ is the mean running time. Respectively p_t is the pupil area at t and $\langle p_t \rangle$ is the mean pupil area over time.

$$z_t = \begin{pmatrix} z_{r,t} \\ z_{p,t} \end{pmatrix} \quad \text{and} \quad z_{r,t} = \frac{r_t - \langle r_t \rangle}{\sigma_r} \quad z_{p,t} = \frac{p_t - \langle p_t \rangle}{\sigma_p} \quad (5.2)$$

behaviour difference: we can extract every representations pairs similarity from the correlation matrix and calculate the difference of the respective behaviour z-vectors, with difference function:

$$\Delta B_{t_1, t_2} = |z_{t_1} - z_{t_2}| = \sqrt{(z_{r,t_1} - z_{r,t_2})^2 + (z_{p,t_1} - z_{p,t_2})^2} \quad (5.3)$$

all sessions

In our analysis we included all available (103) sessions aiming to increase the representativity of the sample and ensure that our findings were robust and generaliseable. Since the sessions differ not only in neural activity and respective similarity of representations, but also in timings of each stimulus replay, we used interpolation to close the gaps between x-values. First we spaced 100 x-values evenly between the minimum and maximum values in the dataset, generating a new set. Subsequently, interpolation was utilised to obtain the respective y-values for each session, allowing to calculate mean representational similarity values between all sessions. At this point it was necessary however to go back into data cleaning and remove some sessions with outlier values, ultimately reducing them to 96 sessions (41 novel sessions and 39 of Vip/Sst mice, where we identified inhibitory cells).

engagement

As proposed by Allen Brain Institute reward rate > 2 , should be considered as engaged mice, where the reward rate is calculated of 25 trial rolling window and provides a measure of the rewards earned per unit time in unit reward/min.

Bibliography

- Deitch, D., Rubin, A., & Ziv, Y. (2021): Representational drift in the mouse visual cortex. *Current biology*, 31(19), 4327–4339.
- Driscoll, L. N., Duncker, L., & Harvey, C. D. (2022): Representational drift: Emerging theories for continual learning and experimental future directions. *Current Opinion in Neurobiology*, 76, 102609.
- Driscoll, L. N., Pettit, N. L., Minderer, M., Chettih, S. N., & Harvey, C. D. (2017): Dynamic reorganization of neuronal activity patterns in parietal cortex. *Cell*, 170(5), 986–999.
- Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009): The mismatch negativity: A review of underlying mechanisms. *Clinical neurophysiology*, 120(3), 453–463.
- Niell, C. M., & Stryker, M. P. (2010): Modulation of visual responses by behavioral state in mouse visual cortex. *Neuron*, 65(4), 472–479.
- Richmond, B. (2013): Neural coding. In D. Jaeger & R. Jung (Eds.), *Encyclopedia of computational neuroscience* (pp. 1–4). Springer New York. https://doi.org/10.1007/978-1-4614-7320-6_398-1
- Rule, M. E., O’Leary, T., & Harvey, C. D. (2019): Causes and consequences of representational drift. *Current opinion in neurobiology*, 58, 141–147.
- Sadeh, S., & Clopath, C. (2022): Contribution of behavioural variability to representational drift. *Elife*, 11, e77907.
- Sheintuch, L., Rubin, A., Brande-Eilat, N., Geva, N., Sadeh, N., Pinchasof, O., & Ziv, Y. (2017): Tracking the same neurons across multiple days in ca2+ imaging data. *Cell reports*, 21(4), 1102–1115.
- Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T. K., Choi, H., Luviano, J. A., et al. (2019): A survey of spiking activity reveals a functional hierarchy of mouse corticothalamic visual areas. *Biorxiv*, 805010.
- Siegle, J. H., Jia, X., Durand, S., Gale, S., Bennett, C., Graddis, N., Heller, G., Ramirez, T. K., Choi, H., Luviano, J. A., et al. (2021): Survey of spiking in the mouse visual system reveals functional hierarchy. *Nature*, 592(7852), 86–92.
- Tang, M. F., Kheradpezhohouh, E., Lee, C. C., Dickinson, J. E., Mattingley, J. B., & Arabzadeh, E. (2023): Expectation violations enhance neuronal encoding of sensory information in mouse primary visual cortex. *Nature Communications*, 14(1), 1196.

picture credits

Figure 2.1, Figure 2.2, Figure 2.3, Figure 2.4 : **Allen Brain Observatory**,
<http://portal.brain-map.org/explore/circuits/visual-behavior-neuropixels>[22.05.23]

Figure 2.5, Figure 2.6: **Allen Brain Observatory**,
<https://portal.brain-map.org/explore/circuits/visual-coding-neuropixels>[22.05.23]

Abbreviations and Symbols

Abbreviations

RS	Representational Similarity
CMOS	Complementary Metal-Oxide-Semiconductor
LGN	lateral geniculate nucleus
LP	lateral posterior nucleus
V1	primary visual cortex
AM	antero-medial area
PM	postero-medial area
AL	antero-lateral area
RL	rostero-lateral area
LM	latero-medial area
CA1	cornu ammonis 1
SDK	Software Development Kit

Symbols

Δ	Difference
\	Excluding Operator

List of Figures

2.1	Visual Change Detection Task	3
2.2	Experimental Session Overview	4
2.3	Image Sets	4
2.4	Inside Task Block	5
2.5	Neuropixel 1.0 Probe	6
2.6	Brain Regions and Optotagging	7
3.1	Image: im111 and Correlation Coefficient Matrix	10
3.2	Single Session - RS over Time in V1	11
3.3	All Sessions - RS over Time in V1	12
3.4	Single Session - RS by Hierarchy: V1, VIS excl. V1 and CA1	13
3.5	All Sessions - RS by Hierarchy: V1, VIS excl. V1 and CA1	14
3.6	Single Session - RS by Cell Type	14
3.7	All Sessions - RS by cell type	15
3.8	Single Session - RS with z-scored Behaviour	16
3.9	Single Session - RS vs. Behaviour Difference with Marginal Distributions	17
3.10	All Sessions - RS vs. Behaviour Difference with Marginal Distributions	18
3.11	Single Session vs. All Sessions - RS by within Active, within Passive	19
3.12	Single Session vs. All Sessions - RS in Relation to Behaviour within Active, within Passive	19
3.13	Single Session - RS for all Images (Familiar)	20
3.14	Single Session - RS for all Images (Novel)	21
3.15	All Sessions - RS by Stimuli Novelty	21
3.16	Single Session - RS, Rewards and Engagement	22
5.1	70 activity vectors with respective stimuli	29