MASTERARBEIT | MASTER'S THESIS

Titel | Title

Optotagging-Based Identification of Confidence Networks in the Orbitofrontal Cortex of Rats

Verfasst von | Submitted by

Charissa-Maree Murphy

Angestrebter akademischer Grad | In partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt |	UA 066 232
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Neuroscience
Betreut von | Supervisor:	Prof. Univ.-Doz. Mag. Dr. Thomas Klausberger

Disclosure

Tools used:

I acknowledge that I have used Grammarly for writing corrections, AutoLatex for writing equations, Google Translate and the help of Eva Gratzl for the German abstract, the Bibcitation extension in Google Docs for help writing the references, and Chat-GPT-4 to help write code for my scripts used in data analysis.

Distribution of labour:

Romana Hauer and Claudia Rudolf helped with training the animals, animal care, and histology. Paul Anderson completed all surgical procedures and wrote many of the scripts used during animal recordings and for subsequent data analysis.

Acknowledgements

With wholehearted gratitude I would like to sincerely thank…

Thomas Klausberger: For the opportunity to conduct research in the lab, it has been an honor to work under your guise. Thank you for all the helpful guidance.

Paul Anderson: My supervisor who had the patience to teach me many of the technical skills I gained during this period. Who laid most of the groundwork for this thesis and carried out many of the technical skills, including all surgical procedures, writing most of the programs used, and designing the new 3D-printed implants. Thank you for all the knowledge you have passed down. I wish you the best for your future academic endeavours and your next exciting chapter; starting a family. No doubt, you will be a great dad!

Romana Hauer: For her pleasant demeanor to work with and for demonstrating how to train and enrich the lives of rats. Romana helped train all of the rats used in this thesis and was responsible for the majority of the animal care.

Mirjiam Lambourne: The previous master's student who spared her free time to transfer on her knowledge and who carved much of the groundwork for me to build my thesis on.

Claudia Rudolph: For your help with training the rats and histology.

Ricardo Merino and Dilara: For teaching me how to section the brains, stain them, mount them, and image them.

Dimitris Mariatos: For being a great office buddy to bounce ideas off of, and for the countless chess games we played- obviously this was extremely conducive to writing a thesis.

The rest of the Klausberger Lab: For the camaraderie and providing a fun place to work.

The Rats: They did the hardest work necessary for this thesis. None of this work would be possible without their sacrifice.

Leo Hutchins: My son, for coming into the lab many weekends with me to help care for the rats. It has been a joy to share my work with you and I hope one day you'll find your own passion.

Jamie Hutchins: Finally but most importantly my loving partner who has supported me every step of the way. It is exceptionally difficult studying whilst being a parent but you have made all of this possible! Thank you for taking on the extra parenting duties so that I could study for exams or spend extra hours at the lab. I know how rare it is, even today, for women to have partners who truly share the load. I'm grateful every day for the freedom to choose this path and for your role in helping make it a reality.

Table of Contents

Declaration **2**
Disclosure **2**
Acknowledgements **3**
Table of Contents **4**
List of Equations **5**
List of Figures **5**
List of Abbreviations **7**
Zusammenfassung **10**
Abstract **11**
Introduction 1 ..... **12**
1.1 Metacognition 12
1.2 Decision Confidence 13
1.3 The Orbital Frontal Cortex 16
1.5 Electrophysiology 18
1.4 Optogenetics 18
1.6 Ramp & Collapse Cells 19
1.6 Rationale 21
1.7 Aim 21
1.8 Hypothesis 21
Methods 2 ..... **22**
2.0 Animals and Housing Environments 22
2.1 Behavioural Training 22
2.2 Anaesthesia and Preparation 24
2.3 Viral injection 24
Cell Type and Viruses Used 24
2.4 Surgical Implants 25
2.5 Histology 25
2.6 Ramp & Collapse Cell Definitions 26
2.7 Optotagging 27
2.8 Data Analysis 27
2.8.1 Spike Sorting 27
2.8.2 Behavioural Analysis: 27
2.8.3 Optotagging analysis: 28
Results 3 ..... **32**
3.0 Behaviour 32
Animal PMA104: 32

Behaviour Across Rats:..... 34
3.1 Histology..... 36
3.2 Optotagging..... 40
3.3 Ramp & Collapse Cells..... 49
3.4 Optotagged Ramp & Collapse Cells..... 50
3.5 Results Summary..... 56
Discussion 4..... 58
4.1 The Story of Ramp and Collapse Cells..... 58
4.2 Decoding Ramp and Collapse Activity..... 59
4.3 The Superior Colliculus..... 62
4.4 The Thalamic Reticular Nucleus..... 63
4.5 The Pontine Nucleus..... 64
4.6 Limitations: to do science is to fail time and time again..... 65
4.7 Future Directions..... 66
4.8 Conclusion..... 70

List of Equations

Equation 1 | *Vevaiometric mean and SEM of waiting time per evidence bin*..... **28**
Equation 2 | *Logistic regression (sigmoid) function for psychometric curve*..... **29**
Equation 3 | *Peak detection window*..... **30**

List of Figures

Figure 1 | *Summary of all the OFC Projection Targets Identified*..... **17**
Figure 2 | *Firing Patterns of R&C Cells During Abandoned & Rewarded Trials*..... **20**
Figure 3 | *Two-Alternative Forced Choice Task*..... **22**
Figure 4 | *Averaged Vevaiometric Plot across Sessions for PMA104*..... **32**
Figure 5 | *Waiting Time Distributions Across Sessions for Rat PMA104*..... **33**
Figure 6 | *Averaged Psychometric Curves Across Sessions for Rat PMA104*..... **33**
Figure 7 | *Averaged Vevaiometric Plot across Sessions for All Animals*..... **34**
Figure 8 | *Waiting Time Distributions Across Sessions for All Animals*..... **35**
Figure 9 | *Averaged Psychometric Curves Across Sessions for All Animals*..... **36**
Figure 10 | *Viral expression and implant locations in Rat PMA98*..... **37**
Figure 11 | *Schematic for Targeted Brain Regions*..... **40**
Figure 12 | *Total Number of Optotagged Units per Projection Area for Rat PMA104*..... **41**
Figure 13 | *Total Number of Optotagged Units & Projection Area Affinty*..... **41**
Figure 14 | *Distribution of Latency, Jitter & Fidelity across Purely Optotagged Units*..... **42**
Figure 15 | *Scatter Matrix of all Optotagged Units that Pass Criteria*..... **43**
Figure 16 | *OFC Optotagged Units by Projection Site in Rat PMA104*..... **44**
Figure 17 | *All Optotagged Units Across Shanks for Rat PMA104*..... **45**
Figure 18 | *Thalamic Reticular Nucleus Optotagged Unit 93 in Rat PMA104*..... **46**
Figure 19 | *TRN & SC Optotagged Unit 219 in the Rat PMA104*..... **47**
Figure 20 | *Unit 251 Optotagged by all Three Projection Sites in PMA104*..... **48**
Figure 21 | *Total Ramp and Collapse Cells Found in Rat PMA104*..... **49**
Figure 22 | *Ramp and Collapse Cell Locations for Rat PMA104*..... **49**
Figure 23 | *Unit 39: Firing Patterns of R&C Cell During Abandoned and Rewarded Trial*..... **50**
Figure 24 | Optotagged *Ramp and Collapse Cell Locations for Rat PMA104*..... **51**
Figure 25 | *Unit 91: Firing Patterns of Optotagged R&C Cell*..... **51**
Figure 26 | *R&C Unit 91: Optotagged by all Three Projection Sites*..... **52**
Figure 27 | *Unit 298: Firing Patterns of Optotagged R&C Cell*..... **53**
Figure 28 | *Ramp & Collapse Unit 298: Optotagged by the TRN in PMA104*..... **53**
Figure 29 | *Unit 102: Firing Patterns of Optotagged R&C Cell*..... **54**
Figure 30 | *Ramp & Collapse Unit 102: Optotagged by the PN and SC in PMA104*..... **55**
Figure 31 | *Network Summary for Optotagged Ramp & Collapse Cells*..... **56**

List of Abbreviations

2AFC	Two Alternative Forced Choice
ACC	Anterior Cingulate Cortex
A/P	Anterior/Posterior
CPP	Centro-parietal positivity
D/V	Dorsal/Ventral
DLEC	Dorsal Entorhinal Cortex
EEG	Electroencephalogram
EC	Entorhinal Cortex
ERP	Event Related Potential
MD	Medial Dorsal Thalamus
OCM	Oculomotor Nucleus
OFC	Orbitofrontal Cortex
PB	Phosphate Buffer
PN	Pontine Nucleus
PFA	Paraformeldehyde
PSTH	Peri-stimulus Time Histogram
R&C	Ramp and Collapse
SALT	Stimulus-Associated Spike Latency Test
SC	Superior Colliculus
Sm	Submedius Nucleus
TBS	Tris-buffered Saline
TRN	Thalamic Reticular Nucleus

This thesis is dedicated to all the women who have been and continue to be denied the chance to pursue education, careers, or dreams of their own. May this be a small reminder that our ambitions matter.

& to my mother who gave up many of her ambitions to raise me,
I am grateful for your sacrifices,
& I hope this makes you proud.

Zusammenfassung

Entscheidungssicherheit ist ein metakognitiver Prozess, der das Vertrauen in eine Wahl widerspiegelt und Verhalten steuert (Liu, 2024). Die Fähigkeit, vergangene Entscheidungen zu bewerten und zukünftige auf Basis aktueller Informationen anzupassen, ist ausschlaggebend für gute Entscheidungsfindung und adaptives Verhalten (Kepecs et al., 2008). Hier wird die Rolle des orbitofrontalen Kortex (OFC) bei der Berechnung von Entscheidungssicherheit in Ratten untersucht. Aufbauend auf früheren Erkenntnissen wurden die axonalen Projektionsziele von OFC-Neuronen identifiziert, die Entscheidungssicherheit repräsentieren, mit besonderem Fokus auf den funktionellen Zelltyp der „Ramp & Collapse Cells“ (R&C-Zellen). Diese Zellen zeigen eine ansteigende Aktivität während der Belohnungserwartung, die kurz vor Entscheidungspunkten abrupt abfällt und somit das Maß der Entscheidungssicherheit widerspiegelt. Während einer Two-Alternative-Forced-Choice-Aufgabe (2AFC) wurde die neuronale Aktivität von Ratten mit Silizium-Sonden aufgezeichnet, gefolgt von antidromen Optotagging, wodurch die räumliche Verteilung optotaggter Zellen, einschließlich der Ramp-&-Collapse-Zellen, sowie ihre Projektionen zum Nucleus reticularis thalami (TRN), Colliculus superior (SC) und Nucleus pontis (PN) sichtbar gemacht wurden. Auf dieser Grundlage konnten die ersten Knotenpunkte des Entscheidungs-Entscheidungssicherheit-Netzwerks skizziert werden. Die Ergebnisse zeigen, dass R&C-Zellen, die Entscheidungssicherheit widerspiegeln, eine klar definierte anatomische und räumliche Spezifität besitzen. Zukünftige Studien sollten darauf abzielen, das vollständige Netzwerk der Entscheidungssicherheit zu kartieren.

Abstract

Decision confidence is a metacognitive process that reflects how certain we feel about a choice and guides how we act on it (Liu, 2024). The ability to reflect on the quality of prior decisions and adjust future choices based on present evidence is crucial for good decision-making and adaptive behaviour (Kepecs *et al.* 2008). This thesis investigates the role of the orbitofrontal cortex (OFC) in encoding decision-confidence in rats. Preliminary work has identified potential regions that the OFC may project to, in continuation of this work this thesis focuses on determining *where* OFC confidence encoding cells project to, with a focus on a functional cell type "Ramp & Collapse Cells" (R&C cells). These cells exhibit ramping activity during reward anticipation and abruptly collapse prior to decision points, reflecting decision confidence. Rats engaged in a two-alternative forced choice task (2AFC) whilst neural activity was recorded with silicon probes. Then antidromic optotagging was performed to identify OFC and R&C projections to subcortical regions. In this thesis I report the spatial distribution of the optotagged cells, including optotagged R&C cells, and where these cells project to; the thalamic reticular nucleus (TRN), superior colliculus (SC) and pontine nucleus (PN). Thus, the first nodes in the decision-confidence network are outlined. This work establishes that confidence-encoding R&C cells have anatomical and spatial specificity. Future work should focus on completing the full confidence encoding network.

Introduction 1

The brain is arguably the most remarkable organ in the body, an organ that defines the story of *who we are,* and is at the core of our best and worst behaviours. Perhaps no story illustrates this better than the curious case of Phineas Gage, a construction worker who in 1848 had a metal tamping iron blast through his prefrontal cortex (PFC), leaving him with changes in personality, temperament and executive function (Harlow, 1999). This landmark case first suggested that the PFC plays a central role in higher-order cognition and behaviour. Gage's famous story has sparked great interest amongst scientists over the years who have tried to unravel the complexities of the PFC, and despite significant strides in this endeavour the full breadth of the area is yet to be established.

Decision-making involves the careful evaluation of benefits and risks to achieve a desired goal, a process that is fundamental to adaptive behaviour in mammals (Morelli, Casagrande & Fort, 2021). The decision-making process enables individuals to weigh multiple factors, such as reward magnitude, effort, and resources, to optimise outcomes. According to Ernst & Paulus (2005), the structure of decision-making can be reduced to four processes: 1.) Input: an individual is presented with several stimuli linked to certain outcomes 2.) Processing: evaluation of the options and the formation of preferences 3.) Output: based on the evaluation of the input, the individual chooses a specific action in alignment with their goals 4.) Feedback: the individual reflects on the action and subsequent outcomes; this reflection guides future decision-making behaviour. Although this framework is reductive, it provides a conceptual basis for examining how dysfunctions at various stages of the decision-making process may manifest in psychiatric conditions. Decision-making deficits are one of the most common symptoms across many psychiatric disorders and can significantly disrupt daily functioning and quality of life, often leading to negative long-term outcomes (Sonuga-Barke *et al.,* 2015). Understanding why and how decisions are made is of the utmost importance for therapeutic intervention.

1.1 Metacognition

Beyond immediate decision-making, many species demonstrate metacognitive abilities: the capacity to identify and reflect on one's own cognitive processes such as perception, thinking, or memory (Fleming & Lau, 2014). The notion that animals possess metacognition is a recent understanding, for which we owe credit to passionate behavioural scientists such as Jane Goodall, a woman who devoted her life to understanding primate behaviour, and for the first time in 1964 documented that wild chimpanzees in the Tanzanian Forest were using tools; good evidence for intelligent behaviour (Goodall, 1964). This discovery inspired researchers to investigate higher-order thinking amongst our beloved neighbors, and with great effort we can

now clearly define that animals such as primates, dolphins, birds, dogs, rats, bees, and more… do indeed possess metacognition (Hampton, 2001; Inman & Shettleworth, 1999; Kepecs *et al*., 2008; Perry & Barron 2013; Smith, 2009; Smith *et al.*, 1995). Thus, metacognitive ability seems to be a matter of degree and not of kind, positioning animals to be excellent models for metacognitive research.

Metacognitive ability supports learning, informs self-awareness, shapes social interactions, and enhances the quality of our decisions, making it a core aspect of human cognition (Frith, 2012; Murphy, Rhodes & Castel, 2024). Despite the importance of metacognition measuring it remains a challenge. According to Rahnev (2025) who evaluated 17 different methods for measuring metacognition, whilst each method was valid, each one may be more or less ideal in different experimental settings, with no single method being perfect. This leaves our understanding of the factors that determine metacognitive ability incomplete, and our current protocols limited due to poor reliability, bias susceptibility, and strong dependence on task performance (Rahnev, 2025). The author states that progress towards improved methodologies in the field requires a more precise evaluation of how metacognitive judgements are formed and applied in decision-making. One major branch of metacognition is decision confidence; our ability to assess how certain we are about our choices, or in other words, how we "know that we know". Whilst humans can report feeling confident or not, a critical question still persists: *how* do individuals assess their confidence in a decision, particularly when confronted with uncertain or imperfect evidence?

1.2 Decision Confidence

The ability to reflect on the quality of prior decisions and adjust future choices based on present evidence is crucial for good decision-making and adaptive behaviour (Kepecs *et al.* 2008). Decision confidence represents a metacognitive process that links subjective certainty to behavioural adjustments (Liu, 2024). For example, if one were asked whether Canada or the United States has the greater landmass, would you feel confident enough in your answer to bet money on it, and how much? And after discovering whether you were right or wrong, how would that outcome affect your willingness to wager more or less money the next time you are given a similar question? In general-knowledge tasks, participants' confidence ratings have been shown to predict their willingness to wager on their answers, illustrating how confidence guides behaviour (Goodie, 2003). Sanders, Hangya & Kepecs (2016) utilised numerous self-ratings to evaluate whether feelings of confidence matched objective measures of correctness, finding that those with higher confidence were more likely to be correct. These findings are supported in academic settings, where students who were conditioned to have increased confidence prior to their exam performed better than those who were conditioned to have reduced confidence, highlighting that confidence, or lack thereof, can influence performance (Moreno, Briñol, & Petty, 2021). On the contrary, overconfidence; an overestimation of one's performance and investment in incorrect beliefs, has serious consequences including political blunders, economic

collapse, war, and mass delusion (Berner & Graber, 2008; Healy & Moore, 2007). In the words of Plous "No problem in judgment and decision making is more prevalent and more potentially catastrophic than overconfidence" (Armstrong & Plous, 1994, *p217*).

Humans are notorious for holding inconsistent beliefs about reality; for example surveys have shown that across all countries the average estimate of immigrants residing in the population is 28% when the actual data is closer to 12%, less than half of the average assumption (Duffy, 2018). So why are humans confident in false measures such as immigration being more prevalent than it is? As Kahneman (2011) explains in his book *Thinking, Fast and Slow*, overestimations are usually driven by cognitive biases, where people form strong beliefs based on limited or emotionally charged information, yet remain unjustifiably confident in their accuracy; an effect he refers to as the "illusion of validity" (Kahneman & Tversky, 1973). Widespread mass delusions amplified by media, social media, echo chambers, political rhetoric, and now AI-generated content, highlight the urgent need to understand how people assess confidence in their beliefs (Lerman Yan, & Wu, 2016; Pierre, 2024; Rodriguez, Bollen & Ahn, 2016). We know that human rationale is compromised largely by biased intuitive heuristics, yet a study demonstrated that humans *do* detect reasoning conflicts even if they choose to follow their gut instincts anyway, with the ability to recognise such bias improving with age (De Neys, Cromheeke, & Osman, 2011). This raises the question of how confidence shapes our reasoning, whether it strengthens or undermines it, and how this influence evolves over time. Improving our understanding of the mechanisms behind decision confidence is pivotal for enhancing the quality of decision-making.

So what is our current understanding of decision confidence? Van Den Berg *et al.* (2016) investigated the relationship between confidence and accuracy in humans, showing that choice, confidence, and reaction time arise from an evidence accumulation process. In this model, the brain integrates sensory information *over time* until a decision threshold is reached, determining both the choice and reaction time, whilst confidence reflects the strength of the accumulated evidence and can continue to update even after the decision is made. In support of this, an EEG study by Ko *et al.* (2024) showed that during decision-making the centro-parietal regions had an increased electrical voltage signal (known as the centro-parietal positivity (CPP)). This signal reflected the accumulation of sensory evidence leading up to a choice: rising more steeply and peaking higher when individuals felt *more* confident, indicating that stronger or more coherent evidence has been integrated. Notably, the CPP can continue to increase even after a decision is made, particularly when confidence is low, supporting the notion that confidence is not fixed at the time of choice but remains flexible and updates when more evidence is accumulated and information processed. Confidence shapes how motor memories and habits are formed; for example, a tennis player who serves with high confidence is more likely to reinforce and retain an effective serving motion than if they perform the serve whilst feeling doubt, despite the outcome of the serve (Nozaki, 2024; Ogasa *et al.*, 2024). Harun *et al.* (2020) used psychophysical reverse correlation and a decision filter model to measure how participants weighted auditory evidence over time when making detection and confidence judgments. They

found that when the task required faster responses, participants relied more on the most recent sounds to make their decisions and rate their confidence levels, showing that people adjust how much evidence they will consider in accordance with how quickly they will need to respond. Further, stress is a known factor to decrease decision confidence (Heereman & Walla, 2011). Indicating that both external pressures, such as time constraints, and internal states such as stress, can shape how evidence is integrated and how certain we feel about our choices.

Persaud *et al.* (2007) introduced a post-decision confidence wagering task where participants were asked to wager based on their confidence; this helped provide measurable indicators of subjective certainty. However, Fleming & Dolan (2010) challenged this notion demonstrating that people's risk preferences influence wagering behaviour, questioning whether post-decision wagering truly reflects metacognitive awareness or is confounded by motivational factors. This underscores the need for reliable neural signatures rather than just confidence ratings. On that note, Fernández-Vargas *et al.* (2021) demonstrated that decision confidence can be decoded from task- and subject-independent EEG signals, with participants who were sure of their decisions showing markedly stronger ERP signatures. These results indicated that neural correlates of confidence are generalisable across individuals and tasks. A reliable marker of decision confidence is reaction time, with high confidence typically reflected in faster responses and low confidence in slower ones (Filevich, Koß & Faivre, 2020; Hellmann, Zehetleitner, & Rausch, 2024). Rahnev, D (2025) showed that confidence and decision speed arise from distinct dynamic networks distributed across cortical layers. They further demonstrated that applying transcranial magnetic stimulation to the dorsolateral prefrontal cortex after a stimulus *increases* confidence without altering accuracy, suggesting that confidence is computed on rapid timescales, sometimes even before a decision is finalised (Xue *et al.,* 2023a). Together these human studies have shown that confidence judgments are rapid, dynamic, and can be modulated, though finding reliable neural signatures still remains a challenge.

Rats have served as an excellent model organism for decision-confidence; however, since the animals cannot self-report, researchers rely on alternative measures, such as temporal investment, to infer confidence. For instance, rodents demonstrate longer waiting times for rewards in tasks where they are more certain of their decisions— a behavioural proxy for confidence (Lak *et al.*, 2014; Masset *et al.*, 2020). Foote and Crystal (2007) showed that rats exhibit metacognition by declining uncertain trials in confidence judgment tasks. When forced to complete trials without this option, their accuracy decreased, indicating the animals can assess when they are likely to be incorrect. Kepecs *et al.* (2008) developed an olfactory two-alternative forced choice (2AFC) task for rats, in which animals discriminated between odour mixtures of varying similarity. They showed that more difficult trials (those with more ambiguous odour cues) led to increased uncertainty, as reflected in shorter post-decision waiting times. During these periods of uncertainty, some OFC neurons increased their firing rate, indicating that they are encoding confidence in proportion to sensory evidence. Joo *et al.* (2021) demonstrated that rats apply confidence judgments to memory: high-confidence recalls led them to wager for larger rewards, whereas low-confidence recalls prompted more cautious choices. This mirrors decision

confidence, in which subjective certainty about a choice directly shapes subsequent strategies and actions. Together, these findings establish rats as a valuable model for studying confidence, showing that they monitor uncertainty and use it to guide behaviour across perception, memory, and decision-making.

In rodents it is clear that the OFC is a central hub for confidence computation; however, there are many unanswered questions. How does the OFC act as a nexus between other brain regions to signal behavioural adjustments, such as how much time to invest waiting or when to update choices based on previous outcomes? What are the local and regional computations for this? To answer these types of questions, a circuit-level understanding of confidence computation is needed. If we could define the network circuitry, we could then observe what happens when a part of the network is modulated, e.g. by excitation or inhibition via optogenetics. Perhaps then we could answer what is necessary or sufficient for confidence evaluation on a global scale.

1.3 The Orbital Frontal Cortex

The orbital frontal cortex (OFC) plays a critical role in decision-making by integrating multiple elements, including reward magnitude, effort, time, and valuation (Fatahi *et al.*, 2018; Fellows, 2011; Schnitzler & Wojtecki, 2012). OFC neurons dynamically encode value signals by linking cues to expected outcomes and updating reward predictions based on new information (Roesch, Taylor & Schoenbaum, 2006; Stolyarova & Izquierdo, 2017). Kepecs *et al.* (2008) identified OFC neurons that increase their firing rates during uncertain or challenging decisions, supporting the hypothesis that these neurons encode choice confidence proportionally to sensory evidence. Temporal wagering tasks further demonstrate the OFC's involvement in decision confidence. The OFC's involvement in temporal cognition, such as in temporal discounting and wagering tasks, highlights its role of integrating time and value to inform decisions (Sosa, Buonomano & Izquierdo, 2021). In these tasks, rodents categorise ambiguous stimuli and report their confidence through the duration they are willing to wait for a reward before aborting the trial. When the OFC is inactivated, this impairs the animal's willingness to wait—a proxy for decision confidence, without affecting decision accuracy (Lak *et al.*, 2014). These findings suggest that the OFC encodes both certainty and the temporal structure of decisions, providing a framework for understanding confidence-driven behaviours. This framework is the very foundation that this thesis aims to build upon.

The OFC operates within an extensive network of brain regions to mediate its role in decision-making and confidence processing. Its bidirectional connections with the amygdala, ventral striatum, and anterior cingulate cortex enable the integration of emotional, motivational, and cognitive information (Rolls, Cheng, Feng, 2020). The amygdala provides inputs related to reward valence and emotional salience, whilst the ventral striatum facilitates reward prediction and learning. The anterior cingulate cortex (ACC), which is involved in conflict monitoring and effort evaluation, works in concert with the OFC to adapt decisions under uncertainty (Fuster,

2015; Schoenbaum *et al.*, 2009). The OFC's projections to sensory and motor areas allow the transformation of value computations into actionable behaviour (Rolls, Cheng, Feng, (2020). This interconnected architecture supports the notion that the OFC not only evaluates reward-based decision variables but also integrates them into a broader framework of decision confidence and temporal investment (Kepecs *et al.,* 2008; Lak *et al.,* 2014; Masset *et al.,* 2020). We know that when confidence is low animals are more likely to be biased by recent reward outcomes, and that by inactivating the OFC this disrupts the reinforcement bias (Lak *et al.*, 2020a). In parallel confidence signals are relayed to dopamine neurons in the ventral tegmental area, allowing adjustment of expectations in response to previous reward outcomes (Lak *et al.*, 2020b). These studies mentioned above, whilst insightful, have only just scratched the surface on OFC confidence networks.

Building on our understanding, structural imaging data from our lab, led by former master's student Mirjam Lambourne, further mapped the connectivity of the OFC to multiple subcortical and cortical regions (Fig.1). Using retrograde viral tracing, seven subcortical targets were identified, including the medial dorsal thalamus (MD), submedius nucleus (Sm), reticular thalamic nucleus (TRN), periaqueductal gray (PAG), oculomotor nucleus (OCM), superior colliculus (SC), and pontine nuclei (Pn), as well as one cortical projection area, the dorsolateral entorhinal cortex (DLEC). These findings were corroborated by either anterograde or retrograde viral injections into each target, confirming broad bidirectional connectivity. What we do not know is where the confidence-encoding neurons project to, leaving room for further investigation.

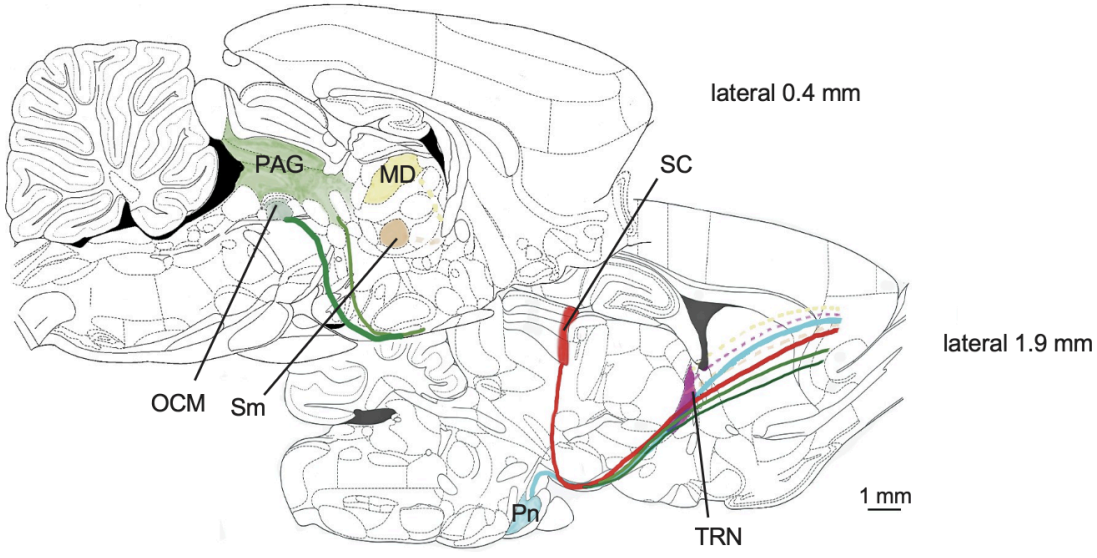

Figure 1 | Summary of all OFC Projection Targets Identified. All target projections are shown, with the exception of the DLEC. Dashed lines highlight only moderate projection evidence. TRN is shown in violet, Sm in orange, PAG in light green, OCM in dark green, and Pn in light blue. PAG,SC, OCM, and Pn have strong connections to the OFC. Background from Paxinos & Watson, 2006. *Figure from Mirjam's Masters thesis, 2024.*

1.5 Electrophysiology

Recent strides in neuroscience have enabled the extensive examination of brains across species, from broad circuits down to cellular single- and multi-neuron recordings, full connectome mapping, transcriptomics, and complete gene profiling (Chen *et al.,* 2019; Kim *et al.,* 2025; White *et al.,* 1986; Verasztó *et al.,* 2020). In recent years, advances in silicon probe technology have significantly increased the number of neurons that can be recorded simultaneously (Buzsáki, 2004). Neuropixel 2.0 probes are implantable four-shank electrodes, with over 5000 sites covering approximately 500-3000 neuronal units, that allow for chronic implantation and enable one to record animals long-term (Steinmetz *et al.*, 2021). These high-density implants are ideal for behavioural recording sessions because they allow for examination of how neurons evolve with experience, learning, or task engagement across prolonged timescales, which leads to the accumulation of rich datasets. The motive behind acquiring such large datasets is to better understand the organisation of the nervous system and how this leads to the emergence of various brain functions (Stevenson & Kording, 2011). Whilst silicon probe development is an impressive feat, neuroscience is still in its infancy, and complete understanding of the whole human brain could take many more decades of research, for silicon probes still only capture a small fraction of the brain's total neuronal population. In reference to technology, a single pyramidal neuron is estimated to be as complex as a five- to eight-layered artificial deep neural network (such AI networks can achieve image classification and natural language processing), now times that by the billions, and we can start to comprehend the scale of complexity of the human nervous system and how we may fall short in capturing its sophistication (Beniaguev, Segev & London*,* 2021). Nonetheless, silicon probes are an advanced tool that allows for detailed insight into the activity of hundreds of neuronal units for prolonged time periods.

1.4 Optogenetics

In a series of fortunate events, researchers created a powerful tool called optogenetics. The origin story of optogenetics started with the discovery that the green algae *Chlamydomonas reinhardti* could swim towards or away from light, a behaviour known as phototaxis (Foster & Smyth, 1980). Later this phototaxis behaviour was found to be mediated by channelrhodopsins, nonspecific ion channels that open in response to blue light which allows positively charged ions to flow into the cell (Nagel *et al.*, 2002). Researchers utilised this knowledge and experimented with ways to engineer neurons to express these light-sensitive channelrhodopsin proteins, allowing ions to flow into neurons and thus elicit action potentials via light stimulation (Banghart *et al.* 2004; Hardie, 2002; Junge, 2002; Takai, 2005; Zemelman 2003). In 2006 Karl Deisseroth was the first to coin the term optogenetics, which he defined as "Optogenetic technology combines genetic targeting of specific neurons or proteins with optical technology for imaging or control of the targets within intact, living neural circuits" (Deisseroth *et al.,* 2006). In other words, optogenetics is a technique that enables researchers to control neural activity using light

stimulation. This is typically achieved through the delivery of light-gated ion channels into the brain via a viral vector containing cell-type specific promoters, resulting in targeted neurons expressing the channels, and enabling them to be activated or inhibited by light (Deisseroth, 2010).

Optogenetics is particularly useful for helping to define the functions of certain neurons and map neural networks. The optotagging method is where neurons that are engineered to express channelrhodopsin are identified by their light-evoked responses during extracellular recording (Lima, Znamenskiy & Zador, 2009). Antidromic optotagging, a lesser-known method, involves stimulating axons in a projection target with light and then recording the spikes that back-propagate to the soma with electrodes (Lima, Znamenskiy & Zador, 2009). Complementing the optotagging technique is the collision test which dates back to classical electrophysiology studies carried out by John Eccles in the 1950s, who electrically stimulated the ventral root of the spine to induce antidromic action potentials that travelled backwards along the axons of motor neurons. To prove the spikes were from the same neuron Eccles used the collision test; if a spontaneous spike had just occurred, and the evoked spike disappeared (or "collided") due to the neuron's brief refractory period, this proved they were from the same cell (Curtis & Andersen, 2007). Stimulus-Associated Spike Latency Test (SALT), an optotagging method used by Kvistani *et al.* (2013), tests whether light stimulation triggers spikes that are time-locked and significantly different from the preceding times (e.g., $p < 0.01$). Units that pass this test are considered “optotagged”; their firing response was evoked by light. Whilst Kvistani’s method was used to identify neurons and their axonal projections, this test could be used to to match such optotagged neurons to their function when combined with electrophysiology recordings during behaviour.

1.6 Ramp & Collapse Cells

Preliminary work in our lab by Lagler *et al.* (*n.d*) has identified a specific functional cell type in the OFC of rats called Ramp and Collapse Cells (R&C). These R&C cells form a distinct subtype of layer 5 pyramidal cells with apical dendrites reaching the frontal pole and axons projecting to subcortical targets, thus positioning them to integrate frontal inputs and convey confidence-related signals downstream. The major functional feature of R&C neurons is that they exhibit ramping activity that rises with reward anticipation and collapses abruptly prior to the decision point to abandon a trial (*Figure 1*). In R&C cells, ramp slope varies with waiting duration; steeper ramps predict shorter waits, whilst shallower ramps predict longer waits, and collapse time closely tracks the decision to abandon, indicating that this activity reflects behavioural confidence. Unlike motor ramping neurons, which peak at movement initiation (Li *et al.*, 2016), R&C cells terminate activity before action onset, indicating that they may play a role in encoding decision confidence, temporal investment, and choice abandonment by integrating this information *over time*. The R&C population is relatively rare in the OFC, accounting for ~1–8% of recorded neurons (Lagler *et al., n.d*), yet from their distinct activity

patterns we can evince that they play a critical role in confidence encoding and choice control. It is unknown *whether* this ramping and collapsing activity is necessary or sufficient for confidence computation, *where* these cells project to, and *how* these cells trigger trial abandonment. Further research is essential for defining how R&C cells contribute to confidence computations and decision-making.

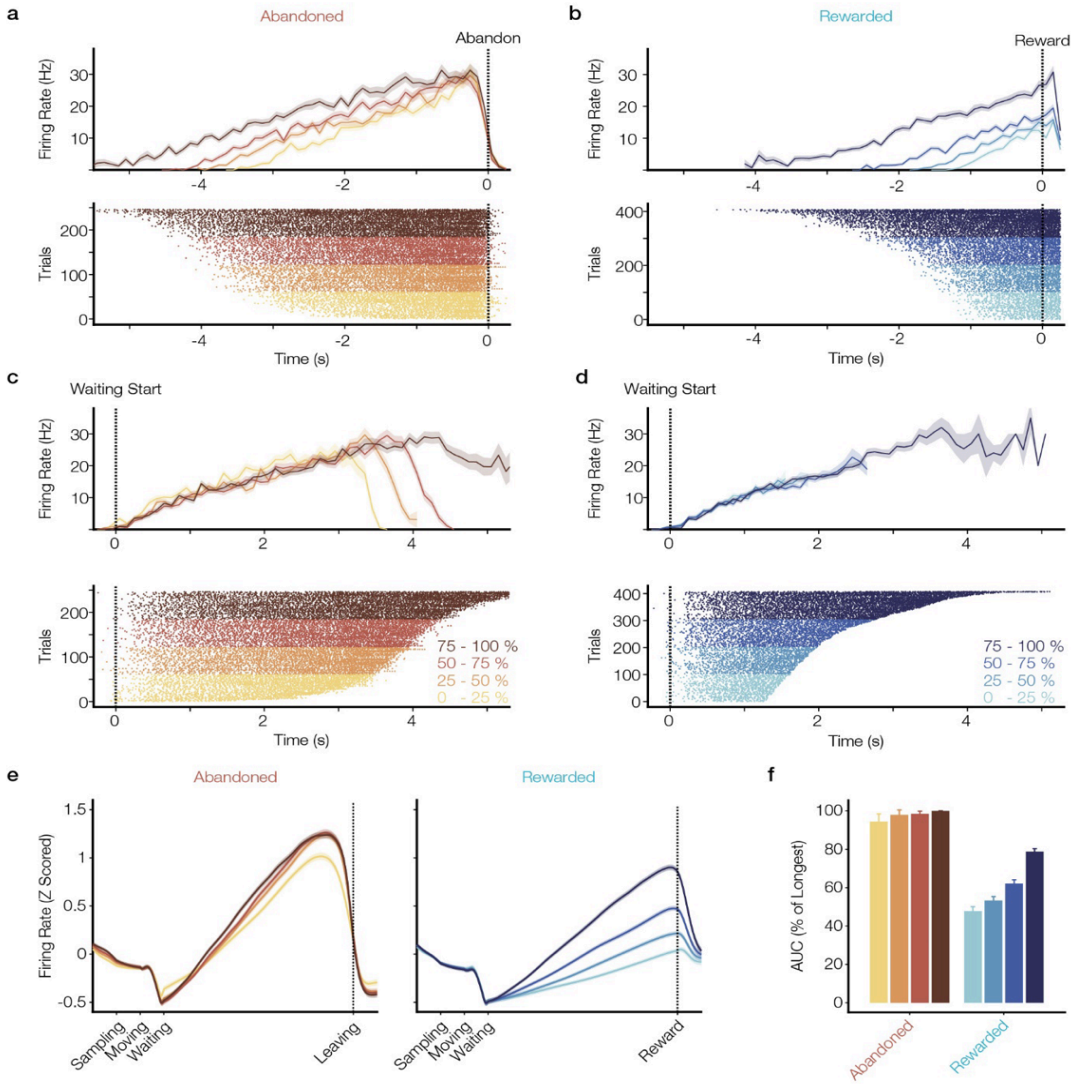

Figure 2 | Firing Patterns of Ramp and Collapse Cells During Abandoned and Rewarded Trials. (a-d) Peri-stimulus time histograms (PSTHs; 100-ms bins, mean ± SEM) and raster plots from a representative ramp/collapse cell recorded with a silicon probe. (a–b) Activity aligned to trial end (abandonment or reward). (c–d) Activity aligned to waiting onset. Colors indicate waiting-time quartiles. (e) Grand averages of all ramp/collapse cells scaled to a uniform time base, showing temporally scaled trajectories in abandoned trials versus interrupted trajectories in rewarded trials. (f) Area under the curve (AUC) of uniform-time PSTHs normalised to the longest quartile, showing consistent trajectories for abandoned trials and truncated trajectories for rewarded trials. Retrieved from Lagler *et al.* (*n.d.*), *unpublished manuscript*.

1.6 Rationale

As discussed, time investment is a behavioural proxy for decision-confidence in rats, with confidence-dependent waiting increasing in proportion to sensory evidence (Kepecs *et al.*, 2008). The novel R&C cells' temporally integrated ramping signals reflect how much time an animal is willing to invest in waiting for a reward, and their collapse profile seems to be signalling when to abandon the task. What is not clear, is how these R&C cells orchestrate such actions within a network. Previous work by Mirjam Lambourne highlighted multiple regions that the OFC projects to: MD, Sm, TRN, PAG, OCM, SC, Pn, and the DLEC. However, what we do not know is w*here R&C* neurons project to. By recording and identifying R&C cells as well as determining their downstream projection sites, we could map the network architecture of R&C cells. In doing so one could then infer what type of signals R&C cells send and receive in order to encode decision confidence.

1.7 Aim

The goal is to record OFC neuronal populations during the 2AFC task in rats and then define *where* the OFC neurons, including Ramp & Collapse cells, project to using the antidromic optotagging method.

1.8 Hypothesis

I hypothesise that functionally defined OFC neurons, including Ramp & Collapse cells, will show a spatial and anatomical specificity in their projection targets.

Methods 2

2.0 Animals and Housing Environments

All experiments involving animals received formal approval and licensing from the Austrian Ministry of Science and the Medical University of Vienna, Austria. This research involved 15 adult male Long Evans rats imported from Charles River in Germany. All rats have a body weight between 300-650 g. The animals were maintained under carefully controlled conditions, with stable temperature and humidity, and a 12-hour alternating light/dark cycle. Recordings were conducted primarily between 8 a.m. and 3 p.m., during our light phase, which corresponds to the rats' dark (active) phase. Sessions took place in a dark room illuminated with red light to maintain the animals' circadian rhythm. Before surgical implantation, the rats were group-housed with no more than four per cage. Post-surgery animals were transitioned to individual housing for their safety. To prevent dust and debris from getting into wounds and/or implants the cages were lined with absorbent bed padding instead of the normal wood shavings that line the cages. Standard rat food pellets were available *ad libitum* and they were given continuous water outside of training/recording periods. The rats were accustomed to handling and spent around 2 hours daily in enriched play environments. During water restriction, Kellogg's Fruit Loops were provided as treats during handling sessions, and cucumber slices were given as a reward at the end of the water restrictive period.

2.1 Behavioural Training

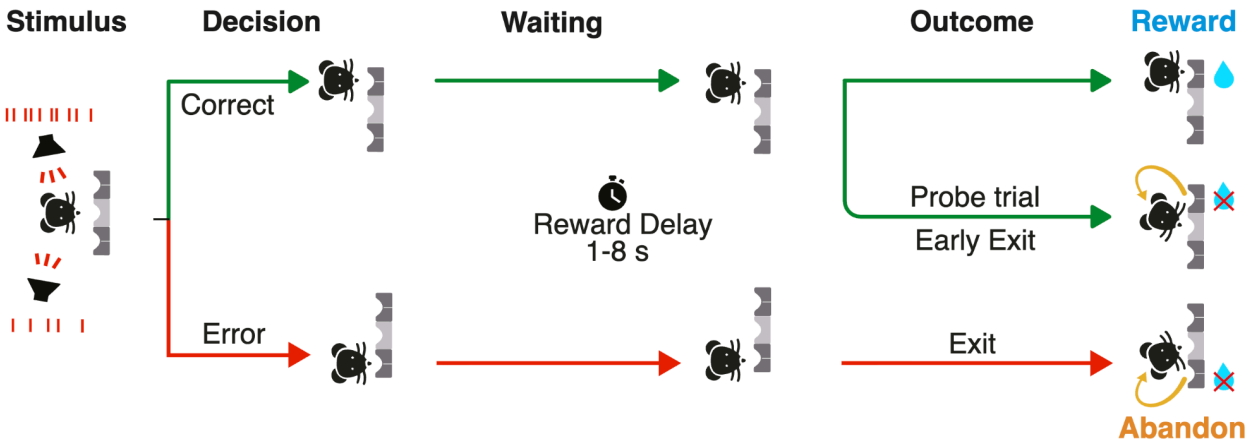

Figure 3 | Two-Alternative Forced Choice Task. The animal is presented with two sound stimuli which prompts the animal to discriminate which speaker played the most clicks (left or right) and subsequently

choose the corresponding port. Outcomes: correct choices = water reward, incorrect choices = no water reward, catch trials = correct choice but no water reward. Figure extracted from preliminary work.

Behavioural procedures were adapted from previous methods by Kepecs *et al.* (2008), Lak *et al.*, (2014), and Masset *et al.* (2020). Naive rats were introduced into the Plexiglass training arena within a sound-dampening box. The training arena is equipped with three ports: a central port activated by a nose poke and two choice ports, left and right, each connected to a water valve for delivering water rewards. The goal of the behavioural training is for the rats to learn to discriminate which speaker is playing more clicks and then choose the correct port, the port on the corresponding side as the speaker (*figure 1*). This behaviour is reinforced with a water reward for correct choices, whilst incorrect choices result in no water reward. On occasion catch trials were introduced, where the animal was correct but was not rewarded; this occurred in no more than 15% of trials. Each port contained photodiodes and phototransistors to detect port entry and exit via disrupted infrared signals. Rats were conditioned to initiate each trial by making a nose poke into the central port (sampling). Once the rats became accustomed to the task, click sounds (a Poisson distributed random click train) were emitted from the left and right speakers simultaneously. The left and right speakers played different quantities of clicks, and the rat was to determine which speaker played more clicks within 20 seconds. The rats indicated their choice by poking their nose into the port that is on the same side as the speaker. The correct side was coupled with a delayed water reward (randomised 2-10 second delay). If the rat left the central port within the first 0.2 seconds of the stimulus, a slightly aversive white noise sound played, followed by a timeout.

Once rats reached ~85% accuracy on easier performance levels, the difficulty level, denoted by alpha (α), was gradually increased. In this task α controls the strength of sensory evidence: lower α values correspond to more trials where the discriminability between left and right is easy, whereas at higher α values the discriminability is harder (values range from 0.1 to 1). When α = 1, the distribution of evidence is flat, meaning that each trial has an equal chance of favouring the left, right, or anywhere in between. When a drawn trial favours neither right or left (equal clicks), the animal is at the decision boundary ($R_L = R_R$ (R = click rate)), where the decision variable (Δ N/N (where ΔN = RR-RL and N = RR+RL)) equals zero. Once the animal's performance was ~85% with α 0.8, catch trials were introduced. In catch trials (5–15% of trials), no reward was provided for either correct or incorrect choices. The time until the rat abandoned the trial was recorded, with a 0.3-second grace period to prevent false detections of abandonment. Each training session lasted 1-3 hours, consisting of 100-1200 trials. Animals had approximately 4-5 training sessions a week for 4-8 weeks until satisfactory performance was achieved. Subsequently, behavioural neural recordings took place for 2-6 weeks. The behavioural measurement systems, Bpod and Pulse Pal were controlled by a custom Matlab (MathWorks, 2023) script to operate the speakers, water valves, photodiodes, and phototransistors. For implanted animals, behaviour was recorded using a Blackfly S BFS-U3-04S2M-CS camera at a frame rate of 400 fps.

2.2 Anaesthesia and Preparation

Rats were anaesthetised using 5% isoflurane in an induction chamber for approximately 5 minutes, followed by maintenance at 3% isoflurane via a gas mask during the surgery. Anaesthetic depth was continuously monitored, and adjustments were made if vital signs, assessed via a pulse oximeter, deviated from normal parameters. The animal was positioned on a thermostatically controlled heating pad to maintain body temperature. The surgical site on the dorsal surface of the skull was shaved and disinfected with iodine. The animal was secured in a stereotaxic frame using ear bars to stabilise the head. Systemic analgesia (meloxicam, 1mg per kg of bodyweight) was injected subcutaneously at the surgical site. Additionally, 2 mL of saline was administered subcutaneously near the hindlimb region to maintain hydration during the procedure.

2.3 Viral injection

Once anaesthesia was stable and animal preparation complete, a midline incision was made along the scalp, extending from the frontal region to the area posterior to the ears. The skin was retracted using surgical retractors to expose the skull. The skull surface was cleaned with a diluted hydrogen peroxide solution applied lightly via a cotton swab, followed by rinsing with sterile saline. Using a stereotaxic frame, bregma and lambda were identified and marked on the skull as anatomical reference points. Target coordinates for the OFC were marked on the skull. These coordinates were guided by the Paxinos & Watson (2005) Rat Brain Atlas. Small 1 mm craniotomies were made at the marked locations using a dental precision drill. A Hamilton syringe mounted on the stereotaxic arm was used to deliver viral vectors into the targeted brain regions.

Cell Type and Viruses Used

Cell Type	Virus
Glutamatergic Cells	rAAV2/5-hSyn-hChR2(H134R)-eYFP (pan-neuronal, but projections are glutamatergic)
GABAergic Cells	pAAV9-Synap-ChETA(E123T/H134R)-eYFP.WPRE.hGH (pan-neuronal, also infects GABAergic cells) Paav9-mdlx-ChR2-mCherry-Fisgell-3 (GABAergic promoter)

***Table 1* | Cell Types Targeted and Respective Viruses Used.** Glutamatergic cells in the OFC (top left) were targeted with the following viruses: Paav9-mdlx-ChR2-mCherry-Fisgell-3 and rAAV2/5-hSyn-hChR2(H134R)-eYFP (top right). GABAergic cells in the OFC (bottom left) were targeted with the pAAV9-Synap-ChETA(E123T/H134R)-eYFP.WPRE.hGH virus (bottom left).

2.4 Surgical Implants

Neuropixel and optic fiber implants

A 1-1.5 mm craniotomy was performed above the OFC coordinates determined by the Paxinos & Watson (2005) rat brain atlas, using stereotaxic measurements referenced from bregma and lambda. A multi-shank Neuropixel 2.0 probe was implanted into the brain tissue using a stereotaxic device. The probe was positioned to the desired depth within the OFC, ensuring alignment with the predetermined coordinates for optimal recording. The probe was mounted onto the drive, then lowered and placed in the target region, and affixed with dental cement. No more than three optic fibres were implanted in any of the regions of interest: orbitofrontal cortex (OFC), superior colliculus (SC), periaqueductal gray (PAG), pontine nucleus (Pn), medial dorsal thalamus (MD), thalamic reticular nucleus (TRN), and the entorhinal cortex (EC).

Each optic fibre was implanted according to the desired target location, using a micromanipulator for precise positioning. Craniotomies, probes and optic fibres were sealed with silicone. The optic fibres were secured to the skull using dental cement, with careful attention to the proper alignment and fixation. Ground screws were implanted into the skull to provide a reference for the electrical signals and were secured with dental cement. A custom 3D printed implant was designed by Paul Anderson; these headplates enabled easier connection and had air holes to prevent condensation buildup and thereby protect the electrical equipment. The headplate was cemented using screws, and the surrounding areas were filled with additional dental cement to ensure stability and prevent any movement. Postoperative care included administering two drops of Novalgin (metamizole sodium) for analgesia. Animals were closely monitored during recovery from anaesthesia to ensure full resumption of normal activity. Two drops of Novalgin were administered every morning for three days post surgery with food. Rat nutrition was modified during this period by providing high calorie gelatinous food, soft food pellets, and rice cereal to minimise the mechanical effort required for food consumption.

2.5 Histology

After three to six weeks of recording (starting minimally 1 week post surgery), the animals were euthanised. Anaesthesia was induced with 5% isoflurane, followed by an overdose of urethane (25% w/v, 1 ml/100 g body weight). Transcardial perfusion was performed with ~200 ml saline, followed by ~200 ml of 4% paraformaldehyde (PFA) containing picric acid and glutaraldehyde. The brains were then extracted, post-fixed in PFA at 4°C overnight, and double rinsed in phosphate buffer (PB). Sagittal brain sections were cut into 100 µm slices using a vibratome and rinsed in PB. Cells were labeled with Hoechst (1:10,000) by adding 1 mL of the dye solution to each well containing a section, followed by incubation at room temperature for 20–30 minutes. Sections were then washed twice with Tris-buffered saline (TBS) and mounted using VectaShield

mounting medium. Images of the mounted sections were captured using a Zeiss Apotome fluorescent microscope and manipulated using the Zeiss Zen Pro interface. A custom-made Matlab script was used for image restitching and background subtraction.

2.6 Ramp & Collapse Cell Definitions

The R&C cell classification was based on their uniform time-base PSTH for leave trials, and all procedures are a copy of methods outlined in Lagler *et al.* (*n.d*). To assess R&C activity across varying waiting times the data was fitted to a uniform time base, akin to methods by Hirokawa *et al.* (2019) and Kobak *et al.* (2016). During the waiting period the firing rates were calculated on a per-trial basis and grouped into 10 ms time bins, then smoothed using a Gaussian kernel with a sigma of 50 ms. To allow for comparisons across trials with different durations, the firing rates were interpolated onto a uniform time base, so that each trial had the same number of time bins, regardless of its actual length. This made it possible to directly compare firing rates of neurons across different length trials. For a cell to be classified as a R&C cell (Ramp Score of 2) the following criteria had to be met: a.) The firing rate had to collapse by ≥ 33% of maximum firing rate *before* trial abandonment. b.) The peak firing rate had to fall within the final 25% of waiting time. c.) The linear correlation over at least 50% of the waiting time must have an R greater than 0.9. Finally, d.) A unit was included only if the increase in firing rate from its minimum value during the first 15% of the waiting period to its peak firing rate exceeded 8 Hz. If all of these criteria passed except one threshold, and that one threshold met a lower threshold of; a.) Firing rate collapse ≥ 25% of the maximum firing rate. b.) Peak firing rate (or 95% of) occurred within the final 33% of the waiting time. c.) R value for a linear correlation with 50% of the firing rate data was greater than 0.7. d.) The absolute increase in firing rate was greater than 5 Hz. Then the cell was classified as having a ramp score of 1.

The analysis was restricted to trials where the waiting times were ≥ 2s and ≤ 8s. The clustered spike times used were from waiting onset to 1s post trial abandonment, and trial times were shifted to negative time values (e.g -8s to -1s) so that the time of trial abandonment was time-locked to 0 seconds. A bin size of 25 ms was used for single-trial spike times and smoothed with a Gaussian kernel sigma of 200 ms. The PSTH trial averaged data excluded timepoints with less than 10 trials, and firing rates were averaged per timepoint. When determining the time of collapse in firing, any sharp discontinuities were filtered out using a Savitzky-Golay with a degree of 3 and a span of 10% of the trial length was applied (Savitzsky, Abraham, & Golay, 1964). The time point where the first derivative of the firing rate crossed 0 was used to identify changes from increasing to decreasing firing rates. The collapse point for a trial or PTSH was defined as the change point where the largest *decrease* in firing rate occurred. Data points that occurred in the first second (or the 15% for uniform time-based data) were not included.

The ramping slope was delineated as; the slope of a linear fit, relative to the minimum value that occurred in the first second (or 15%) of the trial and collapse point. Intrinsic correlations between collapse time, peak firing rate, ramp slope, and trial abandonment were tested against surrogate data. Abandon/leaving time points were recorded via motion tracking using DeepLabCut software on the footage from the Blackfly S BFS-U3-04S2M-CS camera, with a frame rate of 400f ps (Mathis *et al.*, 2018). The abandon/leaving time points were detected via infrared signal changes from the photodiodes and phototransistors.

2.7 Optotagging

Light stimulation was delivered using blue (473 nm) laser diodes housed in an Oxxius laser box controller (Oxxius, France), connected via a fibre optic cable to the implanted optic fibers. Light delivery was controlled through a custom Matlab script developed by Paul Anderson. The optotagging protocol consisted of 2s trials, each containing a specific train of light pulses with different configurations. Pulse duration (1 or 5ms), frequency (2, 5, or 10 Hz), and power (5 or 10 mW, measured at the fiber tip) were combined to create 12 unique light stimulation conditions (2 x 3 x 2). Each condition was repeated 10 times, giving a total of 120 trials. Every trial was followed with a 2s inter-trial interval. To determine whether cells were optotagged I used the SALT method in combination with other criteria (Kvitsiani *et al.,* 2013). For a cell to be considered optotagged the following criterion was applied: peak latency for spikes to occur post stimulus had to be $\leq$16ms, peak jitter $\leq$1ms, peak fidelity $\leq$0.66, SALT and modulation $p \leq 0.05$. Collision Score and ROC thresholds were calculated but not applied. Optotagging plots were generated through a Matlab script made by Paul Anderson. The neuropixel shank's x and y coordinates for each unit were plotted with a horizontal jitter of 12% for better visualisation (so units do not fully overlap when plotting).

2.8 Data Analysis

2.8.1 Spike Sorting

The electrophysiology data from each session was preprocessed and clustered using Kilosort 4.0 and curated manually using Phy to determine noise, multi-unit activity, and well-isolated single units (Pachitariu *et al.*, 2024). For clarification of terminology, a "unit" is the nomenclature that refers to one putative neuron and its activity. The behavioural data, optotagging processing, and R&C classification were quantified using custom Matlab scripts developed by Paul Anderson.

2.8.2 Behavioural Analysis:

The vevaiometric curve (derived from the Greek word βέβαιος, vevaios, which stands for certain) measures how an animal's confidence-dependent waiting time scales with stimulus strength and decision accuracy (Masset *et al.*, 2020). The vevaiometric was calculated by binning trials by evidence, calculating their respective *mean* wait times for correct and error trials, and

plotting them. For each trial the decision variable was signed as $x \in [-1, 1]$ and the waiting time as T (in seconds). Trials that were missing valid waiting time measurements were excluded. The evidence values were divided into seven equal-width bins spanning the range [−1, 1]; where *ck* denotes the center of bin *k*. There were four conditions using the following logical booleans: Correct-Catch (Catch = 1, Correct = 1), Correct-Dropout (Catch = 0, Correct = 1, Rewarded = 0), Error-Catch (Catch = 1, Correct = 0), and Error (non-catch) (Correct = 0). For each condition and evidence bin, all trials with evidence values falling within that bin were grouped together, forming the set *Sk*. The mean waiting time and its standard error of the mean (SEM) were calculated across trials in *Sk*. The mean waiting times and their SEM error bars were plotted against the corresponding bin centers *ck*.

In this analysis the waiting time was used as a behavioural proxy for confidence. Psychometric analyses for behaviour were compared against the sigmoid function, a logistic regression model (Sigmoid Function, *n.d.*). This model takes trial choice/accuracy values and maps them to a probability value between 1 and 0 using the formula:

$$P(\text{right}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \tag{1}$$

Where x = stimulus evidence (number of click train differences is signed so that positive values favour right choice and negative values indicate left choice). The β_0 = bias (shifts the curve left/right to indicate a decision bias) and the β_1 = slope (controls steepness to reflect the sensitivity to the stimulus).

2.8.3 Optotagging analysis:

Trial alignment:

For each recorded unit, spike times were aligned to the onset of every light pulse delivered from the optotagging protocol. This alignment was performed like so: the spike times can be denoted as $S = \{S_1, S_2, \ldots, S_N\}$ and the light onset times as $L = \{\ell_1, \ell_2, \ldots, \ell_M\}$, where *Sj* denotes the time of the *jth* spike and ℓi the onset of the *ith* light pulse. To determine whether spiking activity was time-locked to the light stimulus, spike times were shifted so that each light pulse onset was set to time zero. For each light pulse ℓi, its onset time was subtracted from all spike times. This gave a set of relative spike times: $Si = \{Sj - \ell i\}$, this calculation outputs how many milliseconds before or after the light pulse each spike occurred. Note: *Si* is the list of all spikes, shifted so that the *ith* light pulse starts at time zero.

Time windows and binning:

The time windows for spikes and histogram generation were set to a broad range from −100 to +25 ms around light onset; this range was used so that it included both baseline and early evoked firing. Time windows for the SALT test were −90-0ms for the pre-stimulus period and 2-20 ms for the post-stimulus period. For modulation P, the pre- and post-time windows were −20 to −2 ms and 2 to 20 ms, respectively. The PSTH had a time window of −25 to +50 ms. Spikes were binned in accordance with the sampling interval:

(2)

$$\Delta t = \frac{1}{\text{SampleRate}} \quad \text{and firing rate (Hz)} = \frac{\text{spike count}}{\Delta t}$$

For example, Δt ≈ 0.033 ms at a 30 kHz sampling rate. Firing rates (Hz) were then calculated by dividing the spike counts in each bin by the bin width Δt.

Mean firing rate and baseline calculations:

All spike times from a unit during the optotagging protocol were included (from the first to the last light pulse). The entire period was divided into 1-second bins. The number of spikes in each bin was counted and divided by the bin duration (1 s) to get the firing rate per bin in hertz. The mean firing rate and standard deviation were then calculated across all bins. These values were then used as the baseline firing statistics for subsequent analyses.

PSTH calculations:

For the PSTH window, per-trial binned spike trains were converted to firing rates (Hz) and Gaussian-smoothed (kernel width of ~10 bins). The smoothed traces across all trials were then averaged to get the PSTH for that unit. The PSTHs were the Z-scored using each unit's baseline firing mean and standard deviation. To quantify variability, a 95% confidence interval for the PSTH was calculated at each time point using a t-distribution of the trial means based on the across-trial firing rate variability.

AUC baseline:

The mode of the PSTH$_Z$ (t) was used as the baseline for the area under the curve (AUC). Deviations of the Z-scored PSTH from baseline were quantified as the positive and negative areas under the curve: AUC+ and AUC−.

Peak detection:
Firing peaks from the Z-scored PSTH during the post-stimulus period were based on height and distance thresholds (0.2 and 1 ms, respectively). The peak window was defined as:

(3)

$$t_{\text{pre}} \leq t \leq t_{\text{post}}$$

Here, t_{pre} is the start of the search window relative to light onset, and t_{post} is the end. This equation simply defines the time window in which peaks were allowed to occur. If no peak was detected within this window, fallback cutoffs were estimated using the first-spike latencies from trials with inter-pulse intervals (≥ 110 ms). Only spikes occurring 2–25 ms after light onset were included for this calculation. Cutoffs were set to the median latency ± 3 × SD after outlier removal. Default cutoffs were set to 2 ms (pre) and 15 ms (post), with a minimum pre-window of 2 ms. For each detected peak, the height, time, width, and value were taken directly from the Z-scored PSTH.

Latency:
Latency was defined as the median first spike time after light onset, calculated across all stimulation trials.

Jitter:
Jitter was quantified as the standard deviation of the first-spike latencies (relative to light onset) across all stimulation trials.

Fidelity:
Fidelity was calculated as the fraction of trials containing at least one spike post light stimulation. Units were considered light responsive only if F ≥ 0.6 (spikes occurred in at least 60% of trials).

Note: latency, jitter and fidelity were calculated in seconds.

Results 3

This thesis presents behavioural results from 11 rats during the 2AFC task with a specific focus on rat PMA104, who had successful optotagging (Behaviour, 3.0). An example of histology for rat PMA98 is provided, alongside descriptive statistics for virus injections and implants of all animals involved in experiments (Histology, 3.1). I describe the optotagging results and statistics (Optotagging, 3.2) followed by an account of the R&C cells found with corresponding optotagging status (Ramp & Collapse Cells, 3.3).

3.0 Behaviour

Animal PMA104:

For animal PMA104 there were 24 2AFC behavioural sessions comprising 12,798 trials in total that were used for this analysis. Sessions with less than 300 trials were excluded. These excluded sessions were typically from the first recording periods, where the animal was still getting used to the recording setup or they were from Monday recordings post 72hr water access, making the animal unmotivated to do the task. PMA104 had a performance accuracy (probability of being correct on any given trial) of ~80%. In total, ~12% of sampling dropouts (when the animal fails to activate the middle port for longer than 0.2s, resulting in insufficient evidence accumulation) occurred across sessions. A ~5% real drop out rate (when a rat starts waiting but drops out before reward delivery) occurred across all trials. The performance for PMA104 is shown below.

2AFC Average Performance for Animal PMA104

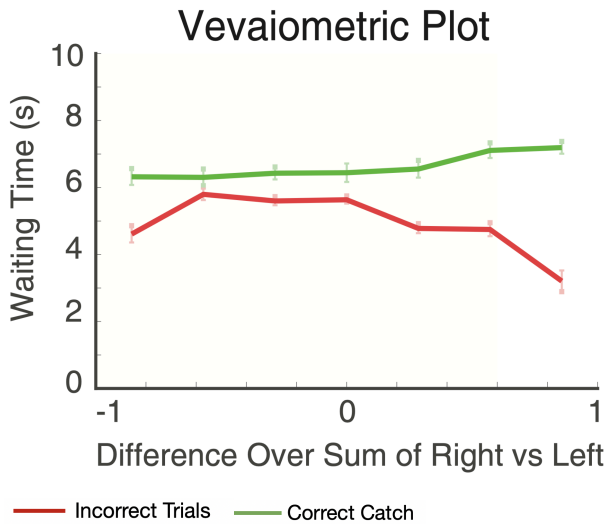

Figure 4 | Averaged Vevaiometric Plot across Sessions for PMA104.
Plot showing mean waiting time as a function of stimulus strength, with catch trials in green and incorrect trials in red. For the difference over sum, -1 = strong evidence for the left, 0 = equal evidence for both sides, and 1 = strong evidence for the right. Correct catch = 386, correct drop out = 91, error catch = 82, error = 2057.

The vevaiometric curve (Fig.4) demonstrates an X-shaped trial separation: when evidence is weak correct and error trials meet near the center time bin; as sensory evidence increased, waiting times diverged to the extremes, with correct catch trial wait times falling between ~6–7 s, and incorrect waiting times being shorter, around ~3–5.5 s. This supports that waiting time serves as a behavioural proxy for decision confidence (Kepecs *et al.*, 2008; Lak *et al.*, 2014; Masset *et al.*, 2020).

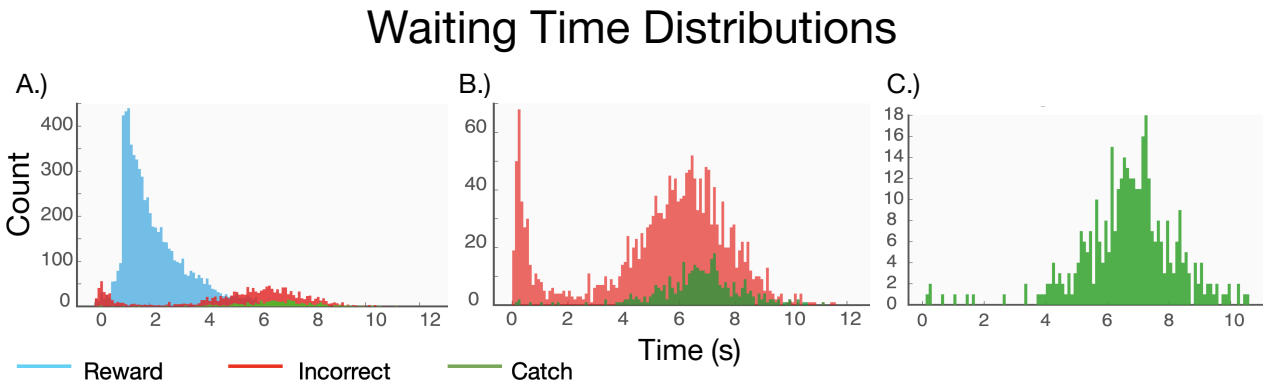

Figure 5 | Waiting Time Distributions Across Sessions for Rat PMA104. Distribution of wait times for incorrect (red), reward (blue), and catch (green) trials. A.) All trials shown: correct, incorrect and catch. B.) Incorrect vs catch. C.) Catch trials only.

Distributions of waiting times across trial outcomes demonstrate the confidence-dependent structure of temporal investment. Correct reward trials terminated earlier due to reward interruption, with wait times peaking around 2–4s (Fig.5A). Incorrect trials had shorter waiting durations compared to catch trials, with incorrect trial peaks concentrated around 0–1s and 5–7s (Fig.5B). Catch trials had the longest wait times, with trial peaks clustered at 6–8 s (Fig.5C). These results reflect that time investment was proportional to predicted internal measures of decision confidence.

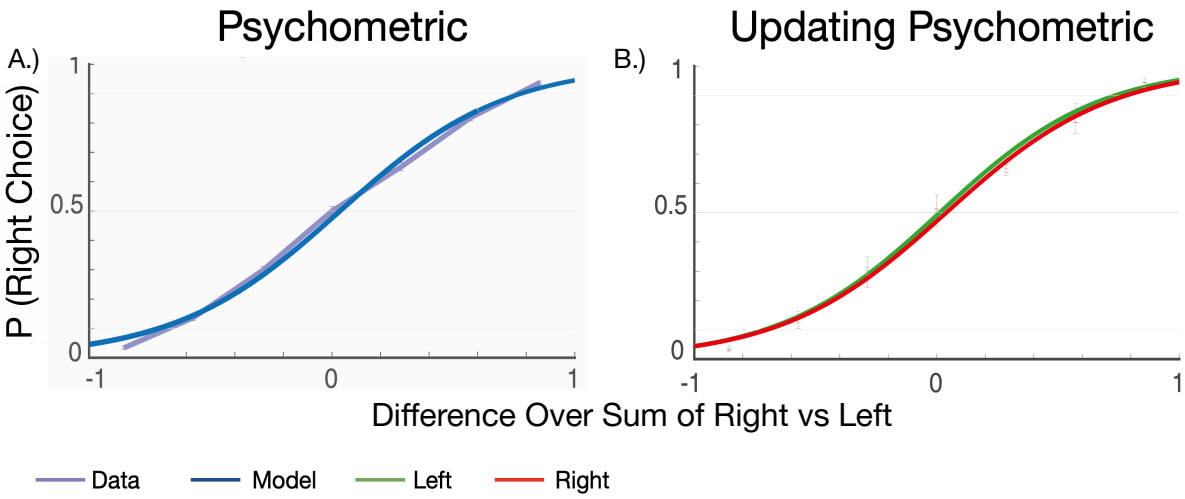

Figure 6 | Averaged Psychometric Curves Across Sessions for Rat PMA104. A.) Psychometric data (purple) shows the probability of a right choice relative to the stimulus strength, plotted against the model fit (blue). B.) The updating psychometric function shows the probability of making a right choice (red) and a left choice (green), this is compared with the model psychometric curve (purple). For the psychometric curve slope = 2.922, midpoint = 0.043, goodness of fit = 0.004. For the difference over sum, -1 = strong evidence for the left, 0 = equal evidence for both sides, and 1 = strong evidence for the right. *Note:* The model fits can not be seen very clearly due to curved fits overlapping.

The psychometric curve (Fig.6A) demonstrates that perceptual evidence is strongly guiding the animal's choices; as the difference over sum (DOS) increased, so did the probability of selecting the corresponding side. Thus, PMA104 had a good evidence to choice relationship. This is consistent with evidence-based decision-making observed in rodent discrimination tasks (Kepecs *et al*., 2008; Kepecs & Mainen, 2012). The updating psychometric plot was almost perfectly matched to the model fit (Fig.6B). This is further proof that PMA104 used accumulated evidence to guide their choices, with minimal decision bias. Together the vevaiometric, wait time distributions and psychometric results (figs 4, 5 & 6, respectively) validate that the animal PMA104 possessed confident-dependent waiting behaviour.

Behaviour Across Rats:

For the behavioural data from the 2AFC task, recording sessions across 11 rats were pooled for this analysis. These 11 animals were: PMA89, PMA92, PMA94, PMA95, PMA96, PMA98, PMA99, PMA100, PMA101, PMA104 and PMA105. In total ~44, 557 trials were evaluated across animals. The behaviour of all rats is shown below.

2AFC Average Performance Across All Animals and Sessions

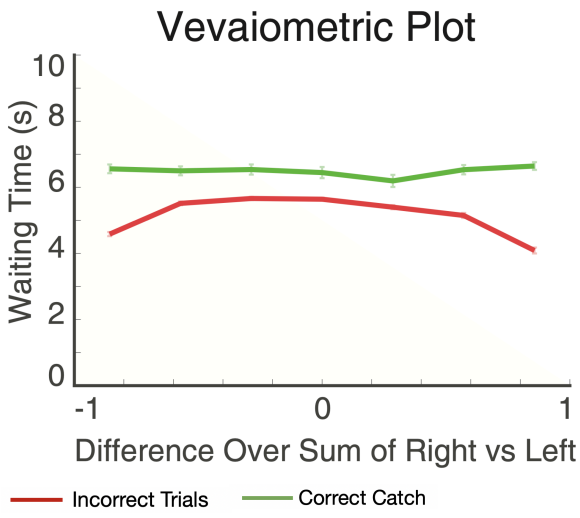

Figure 7 | Averaged Vevaiometric Plot across Sessions for All Animals.
Plot showing mean waiting time as a function of stimulus strength, with catch trials in green and incorrect trials in red. For the difference over sum, -1 = strong evidence for the left, 0 = equal evidence for both sides, and 1 = strong evidence for the right. Correct catch = 966, correct drop out = 1296, error catch = 272, error = 8712.

Akin to previous findings by PMA104 (Fig.4) the vevaiometric curve across all animal sessions (Fig.7) has a X-shaped trial separation: when evidence is weak, correct and error trials meet near the center time bin; as sensory evidence increased, waiting times diverged to the extremes, with correct catch waiting times somewhere between ~6–7s, and incorrect waiting times being shorter around ~4–5.5s. This population data also shows, once more, that these findings are in line with prior demonstrations that waiting time serves as a behavioural proxy for decision confidence (Kepecs *et al.*, 2008; Lak *et al.*, 2014; Masset *et al.*, 2020).

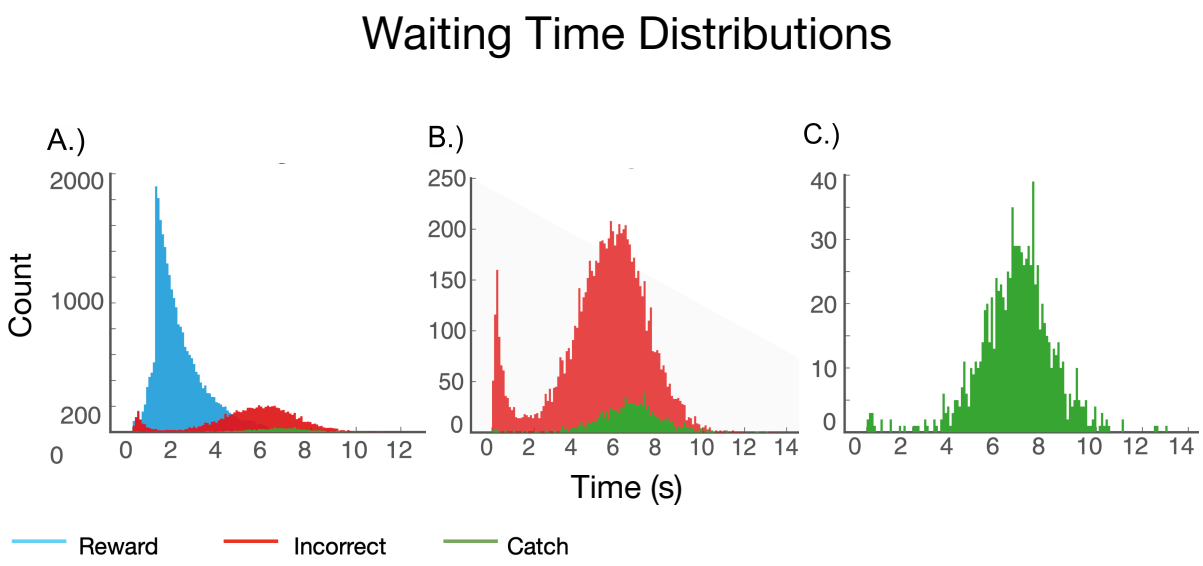

Figure 8 | Waiting Time Distributions Across Sessions for All Animals. Distribution of wait times for incorrect (red), reward (blue), and catch (green) trials. A.) All trials shown: correct, incorrect and catch. B.) Incorrect vs catch. C.) Catch trials only.

Parallel to findings by PMA104 (Fig.5) the distributions of waiting times across different trial outcomes reflected a confidence-dependent pattern for temporal investment. As expected, correct reward trials terminated earlier due to reward interruption, with wait times centering around 1–3s (Fig.8A). Incorrect trials across animals had shorter waiting durations compared to catch trials, with the peaks clustering around 0–1s and 4–8s (Fig.8B). Incorrect waiting durations were slightly longer compared to the single animal PMA104 (Fig.5B). Showing variance at the population level. Catch trials had the longest wait times, with trials clustered at 6–8 s (Fig.8C).

These results attest that time investment across all animals was proportional to internal measures of decision confidence. The catch trials showed greater persistence, with the animals waiting longer for reward, whereas error trials had higher levels of disengagement, especially during the earlier epochs (0-1s peaks). These wait time distributions demonstrate that, at the population level, waiting strategies scale in correspondence with subjective certainty.

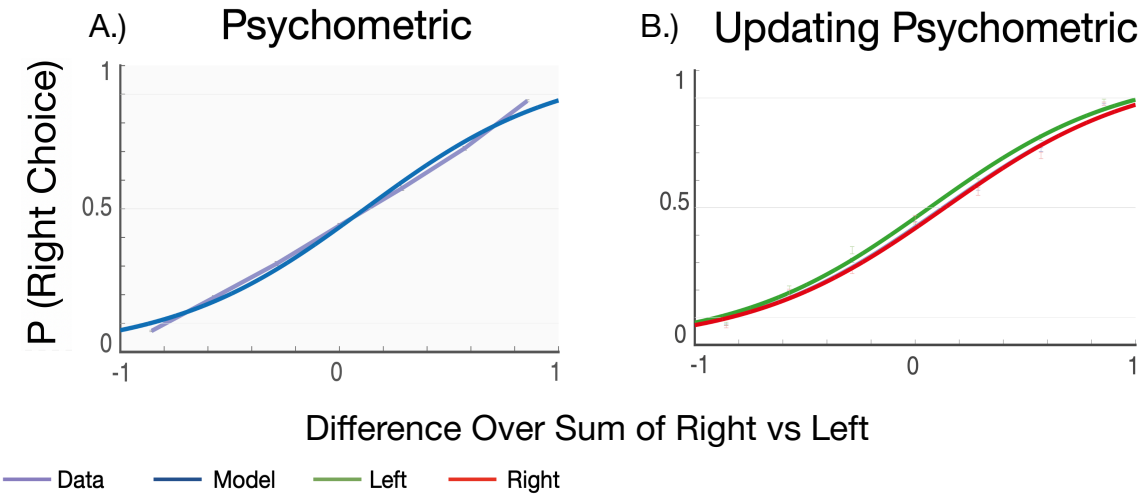

Figure 9 | Averaged Psychometric Curves Across Sessions for All Animals. A.) Psychometric data (purple) showing the probability of a right choice relative to the stimulus strength, plotted against the model fit (blue). B.) The updating psychometric function shows the probability of making a right choice (red) and a left choice (green), this is compared with the model psychometric curve (purple). For the psychometric curve slope = 2.239, midpoint = 0.116, goodness of fit = 0.004. For the difference over sum, -1 = strong evidence for the left, 0 = equal evidence for both sides, and 1 = strong evidence for the right. *Note:* The model fits can not be seen very clearly due to curved fits overlapping.

Across animals, psychometric curves showed that perceptual evidence influenced the animals' behaviour; as the difference over sum (DOS) increased, so did the probability of selecting the corresponding side. However, compared to the single animal PMA104 (Fig.6A), the population psychometric curves had a slightly reduced sigmoid shape. This suggests that some animals may have had diminished sensitivity to the stimulus strength. Additionally, the updating psychometric reflects that over the population there was a slightly higher probability of the animals choosing left if the previous trial that occurred was on the left, indicating a slight choice bias for the left side (Fig.9B). Despite this slight variance, the vevaiometric, wait time distributions and psychometric behavioural results from animal PMA104 and the population validate that the rats possess confident-dependent waiting behaviour. This is in alignment with previous research (Kepecs *et al*., 2008; Kepecs & Mainen, 2012)

3.1 Histology

The purpose of the histology is to identify virus expression and implant locations. 13 animals have been perfused, sectioned, imaged, and analysed. There are 11 animals with implants and virus expression in the correct locations. The remaining animals are still being recorded or were used for non-recoverable optotagging experiments, only PMA104 had successful optotagging (*Table 1*). Figure 10 shows one example of histology identification in the rat PMA98.

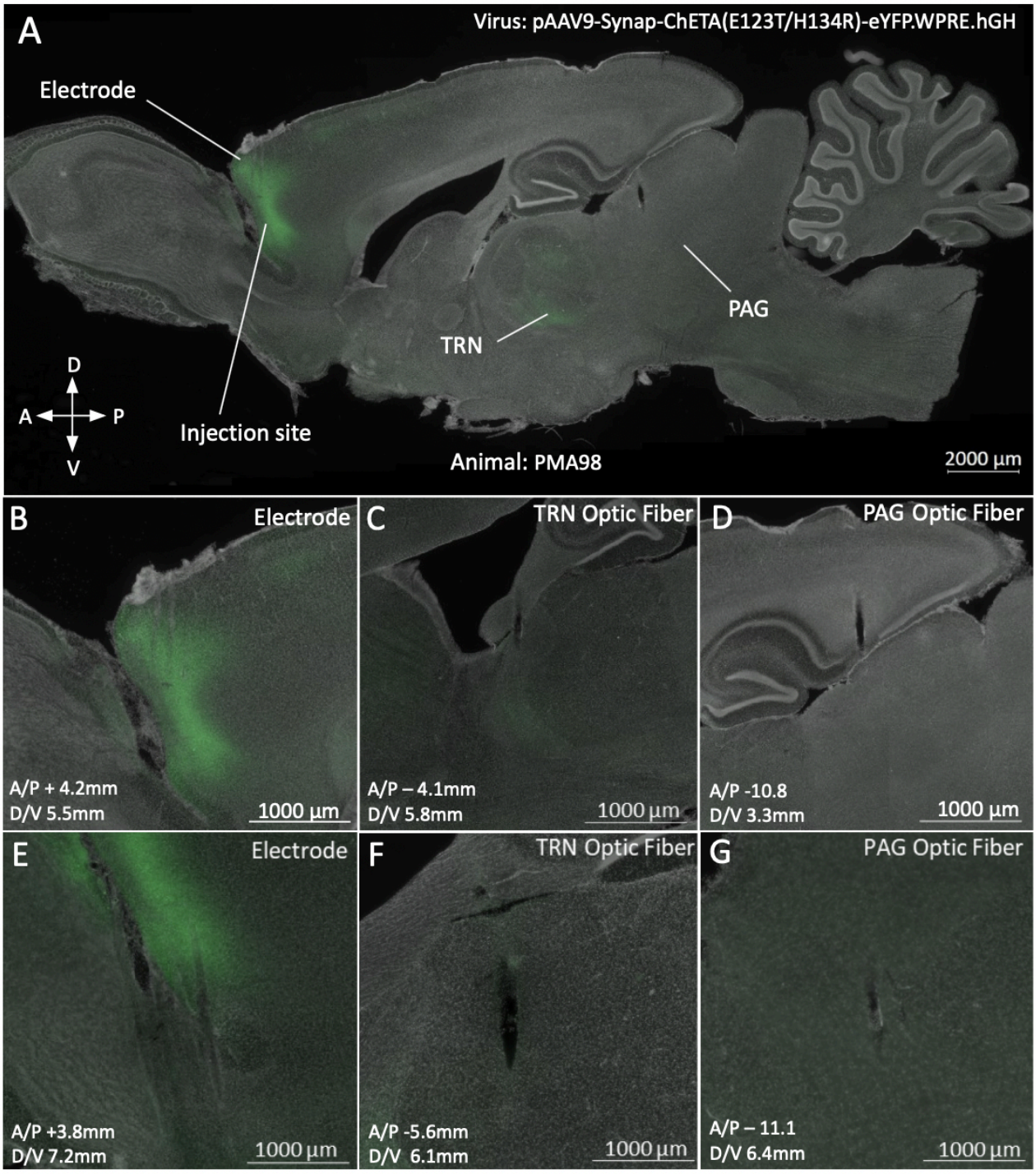

Figure 10 | Viral expression and implant locations in Rat PMA98. (A) Sagittal section showing viral expression (green) in animal PMA98. Electrode, injection site, thalamic reticular nucleus (TRN), and periaqueductal gray (PAG) are presented. (B, E) Coronal sections through the injection site and electrode tracks at anterior–posterior (A/P) +4.2 mm, depth (D/V) 5.5 mm (B) and A/P +3.8 mm, D/V 7.2 mm (E). (C, F) Coronal sections showing TRN optic fiber placement at A/P –4.1 mm, D/V 5.8 mm (C) and A/P –5.6 mm, D/V 6.1 mm (F). (D, G) Coronal sections showing PAG optic fiber placement at A/P –10.8 mm, D/V 3.3 mm (D) and A/P –11.1 mm, D/V 6.4 mm (G). Scale bars: 2000 µm (A), 1000 µm (B–G).

Animals	Virus	Virus Target	Injection Volume	Titre (GC/ml) *x 10^13*	A/P M/L (mm)	D/V (mm)	Electrode Target	A/P M/L (mm)	Electrode Identified	Fiber Targets	A/P (mm)	M/L (mm)	Fibers Identified
PMA89	pAAV9-Synap-ChETA(E123T/H134R)-eYFP.WPRE.Hgh	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	20.46 47.6	Yes	MD SC Pn	11.76 8.06 6.96	49.37 51.0 48.6	Yes
PMA90	Paav9-mdlx-ChR2-mCherry-Fisgell-3	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	18 49.83	Yes	OFC	18.2	50.93	Yes
PMA92	pAAV9-Synap-ChETA(E123T/H134R)-eYFP.WPRE.hGH	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	14.96 47.6	Yes	TRN EC PAG	12.95 10.65 8.45	51.92 51.16 50.45	Yes
PMA94	pAAV9-Synap-ChETA(E123T/H134R)-eYFP.WPRE.hGH	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	18.21 50.61	Yes	TRN PAG	7.71 12.21	50.50 51.97	Yes
PMA95	pAAV9-Synap-ChETA(E123T/H134R)-eYFP.WPRE.hGH	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	15.47 49.6	Yes	MD SC	8.27 4.57	48.93 50.56	Yes
PMA96	Paav9-mdlx-ChR2-mCherry-Fisgell-3	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	16.34 49.6	Yes	OFC	17.48	35.54	Yes
PMA97	Paav9-mdlx-ChR2-mCherry-Fisgell-3	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	15.12 49.9	Yes	OFC	15.12	49.36	Yes

Animals	Virus	Virus Target	Injection Volume	Titre (GC/ml) *x 10^13*	A/P M/L (mm)	D/V (mm)	Electrode Target	A/P M/L (mm)	Electrode Identified	Fiber Targets	A/P (mm)	M/L (mm)	Fiber Identified
PMA98	pAAV9-Synap-ChETA(E123T/H134R) -eYFP.WPRE.hGH	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	17.15 51.25	Yes	TRN PAG	6.65 11.15	51:4 52:6	Yes
PMA99	pAAV9-Synap-ChETA(E123T/H134R) -eYFP.WPRE.hGH	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	14.12 49.55	Yes	MD SC	6.92 3.22	49.02 50.65	Yes
PMA100	pAAV9-Synap-ChETA(E123T/H134R) -eYFP.WPRE.hGH	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	16.24 50.38	TBD	TRN PAG	5.74 10.24	50.27 51.74	TBD
PMA101	pAAV9-Synap-ChETA(E123T/H134R) -eYFP.WPRE.hGH	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	15.33 49.02	Yes	EC Pn	1.61 3.33	51.07 48.02	Yes
PMA102	Paav9-mdlx-ChR2-mCherry-Fisgell-3	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	13.4 51.48	Yes	OFC	4.3	1.8	Yes
PMA103	pAAV9-Synap-ChETA(E123T/H134R) -eYFP.WPRE.hGH rAAV2/5-hSyn-hChR2(H134R)-eYFP	OFC- Left OFC-Right	0.25nl	2.4	4.7 1.8	2.3 (3.5)	OFC	15.78 51.93	Yes	MD TRN PAG Pn	8.58 9.78 5.28 3.78	51.2 46.33 46.03 46.64	Yes
PMA104	rAAV2/5-hSyn-hChR2(H134R)-eYFP (Successful optotagging)	OFC	100nl	2.4	4.7 1.8	2.3 (3.5)	OFC	15.38 50.37	Still Recording	TRN SC Pn	9.38 4.48 3.38	51.71 51.89 49.57	TBD

Table 2 | Virus and Surgical Implant Summary. For each animal, the table lists the viral construct, injection parameters (volume, titre), and injection coordinates (AP/ML/DV), as well as electrode and optic-fiber targets with their stereotaxic coordinates. Identified = tracks were confirmed via histology within the intended target (approximate). Yes = verified implant seen in histology images; TBD = verification pending; Still recording = recording ongoing; No = implant/virus not observed. Abbreviations: OFC, orbitofrontal cortex; TRN, thalamic reticular nucleus; MD, mediodorsal thalamus; SC, superior colliculus; PAG, periaqueductal gray; EC, entorhinal cortex; Pn, pontine nuclei; AP/ML/DV, anteroposterior/mediolateral/dorsoventral (relative to bregma, *Rat Brain Atlas*).

3.2 Optotagging

The optotagging was successful for animal PMA104 only. In total 24 behavioural recording sessions were selected for the analysis of PMA104. For the optotagging there were a total of 37 sessions, comprising 6,014 trials. PMA104 had three fiber implants, one per region in the PN, SC, and TRN. All data presented in this section is from PMA104.

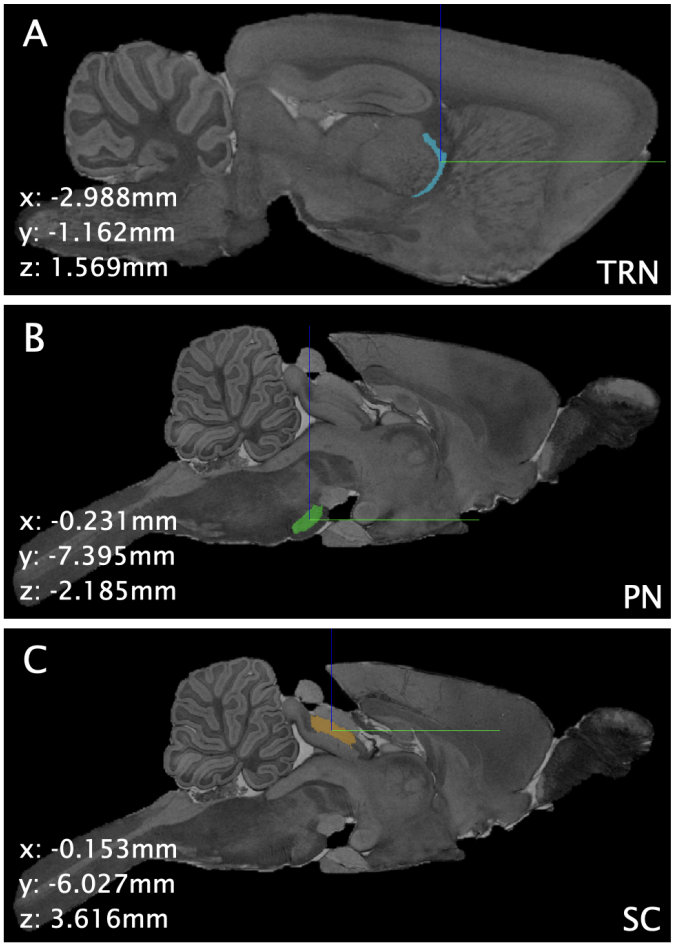

Figure 11 | Schematic for Targeted Brain Regions. Shown are high resolution magnetic resonance images mapping the three brain regions that this study aimed to target with light via optic fibre implants in Rat PMA104. The intersection of the blue and green lines mark the reference points for the x and y coordinates taken from the atlas. A.) location of the thalamic reticular nucleus (blue). B.) location of the pontine nucleus (green). C.) location of the superior colliculus (yellow).

These images are extracted from the Waxholm Space Atlas of the Sprague Dawley Rat Brain (Papp *et al.,* 2014).

Note: This figure is for diagrammatic purposes only. Coordinates shown on the figures are ***not*** the same as the stereotaxic coordinates used for the optic fibre implants (*for optic fibre implant coordinates please see Table 2*).

A total of 184 purely optotagged units were found: with 47 in the PN, 76 in the SC and 96 in the TRN (Fig.12). A substantial number of units were "potentially" optotagged, meaning that these units passed the majority of optotagging criteria, except for peak fidelity and modP measures (*See Methods*). A total of 1638 units were potentially optotagged with 561 in the PN, 588 in the SC and 1119 in the TRN. The majority of units (6993) recorded were not optotagged.

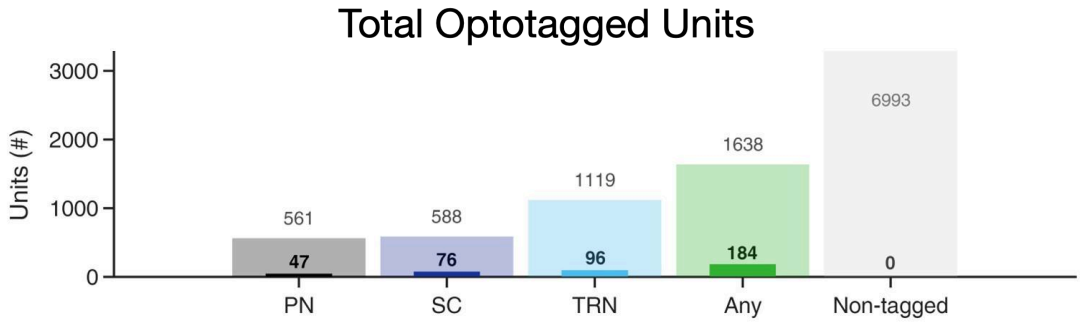

Figure 12 | Total Number of Optotagged Units per Projection Area for Rat PMA104. Shown are the number of units optotagged by the projection targets: PN (black), SC (dark blue), TRN (light blue), the total of all optotagged units (green), and non-optotagged units (light grey). Darker shades represent purely optotagged cells (they pass all criteria outlined in *Methods*) and light shaded colors are the total of "potentially" optotagged cells (cells that passed most criteria but not all). *Note:* The numbers from PN, SC and TRN do not add up in "Any" because the number of units overlap e.g a unit may project to both the TRN and SC and is only counted as one unit not two. *Note*: Figure is cropped for scaling aesthetics and the total number is written by each bar in the plot.

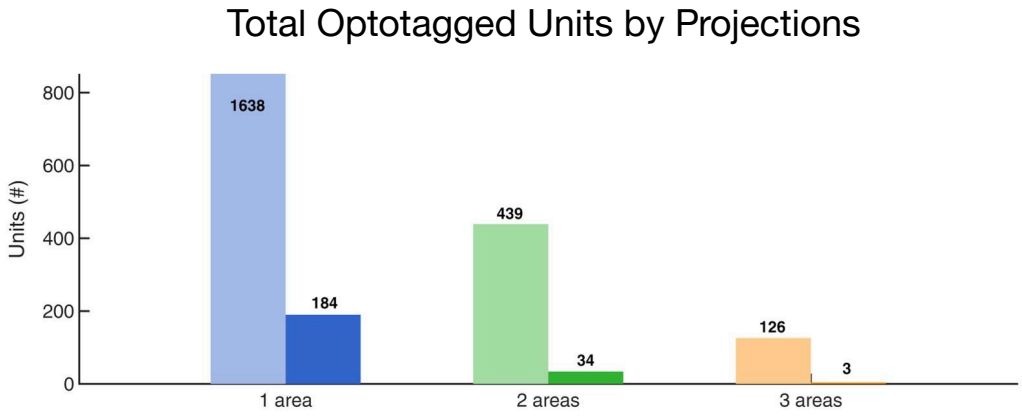

Figure 13 | Total Number of Optotagged Units & Projection Area Affinity for Rat PMA104. Shown are the number of units optotagged and their respective projection affinities: 1 area = number of units optotagged by one brain region only (blue), 2 areas = number of units optotagged by two brain regions (green), 3 = number of units optotagged by all three regions (PN, SC &TRN) in orange. Light shades represent "potentially optotagged" units (cells that passed most criteria but not all, *see Methods*) and darker shades are the purely optotagged units. Note: The figure has cropped the 1 area bar plot (light blue) in half for scaling aesthetics, and the total number for this bar plot is written inside the bar.

In Figure 13, a total of 184 units were purely optotagged by a single region (PN, SC, or TRN), and 1,638 units were classified as potentially optotagged by one region. An additional 34 purely optotagged and 439 potentially optotagged units responded to stimulation from two regions, whilst 3 purely optotagged and 126 potentially optotagged units responded to light activation in all three regions. Across the entire dataset of 6,993 recorded units, only a small fraction were purely optotagged: approximately 2.6% by one region, 0.5% by two regions, and 0.04% by all three regions. Amongst potentially optotagged cells, 23.4% were associated with one region, 6.3% with two regions, and 1.8% with all three regions. Ergo, it is rare for units to project to multiple regions, especially all three.

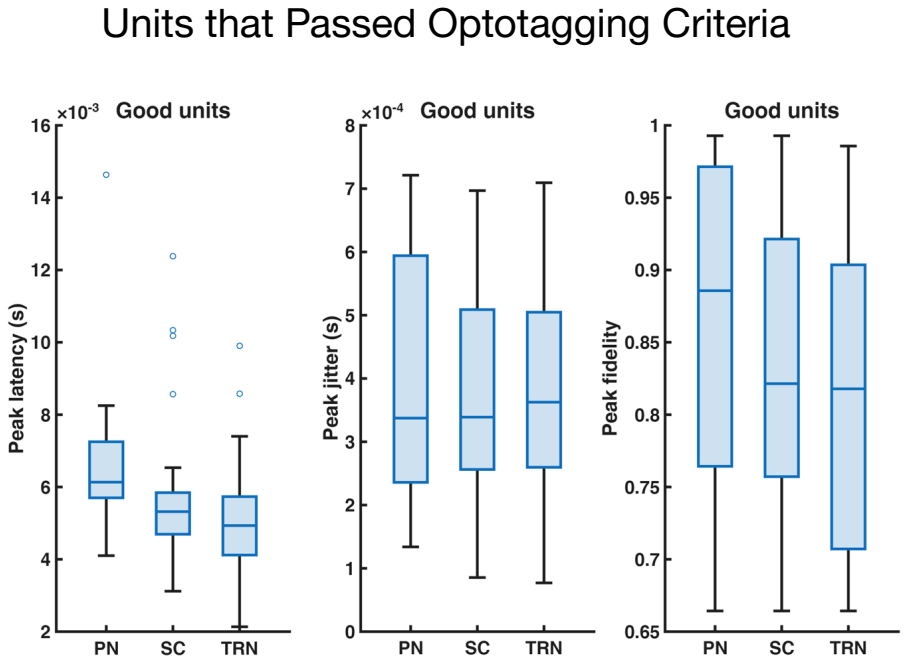

Figure 14 | Distribution of Latency, Jitter & Fidelity across Purely Optotagged Units. Boxplots show the distribution of latency, jitter, and fidelity for all optotagged units recorded in the PN, SC, and TRN. Median values are indicated by horizontal lines within each box, with boxes representing interquartile ranges and whiskers showing the full data range (excluding outliers, shown as circles). Good units = pure optotagged units.

TRN neurons had the shortest response latencies, whereas PN and SC neurons had slightly longer latencies which could be due to their greater physical distance from the OFC. Peak jitter and fidelity were similar across regions. Thus, optotagging was consistent across all three regions, with only minimal regional differences in response timing.

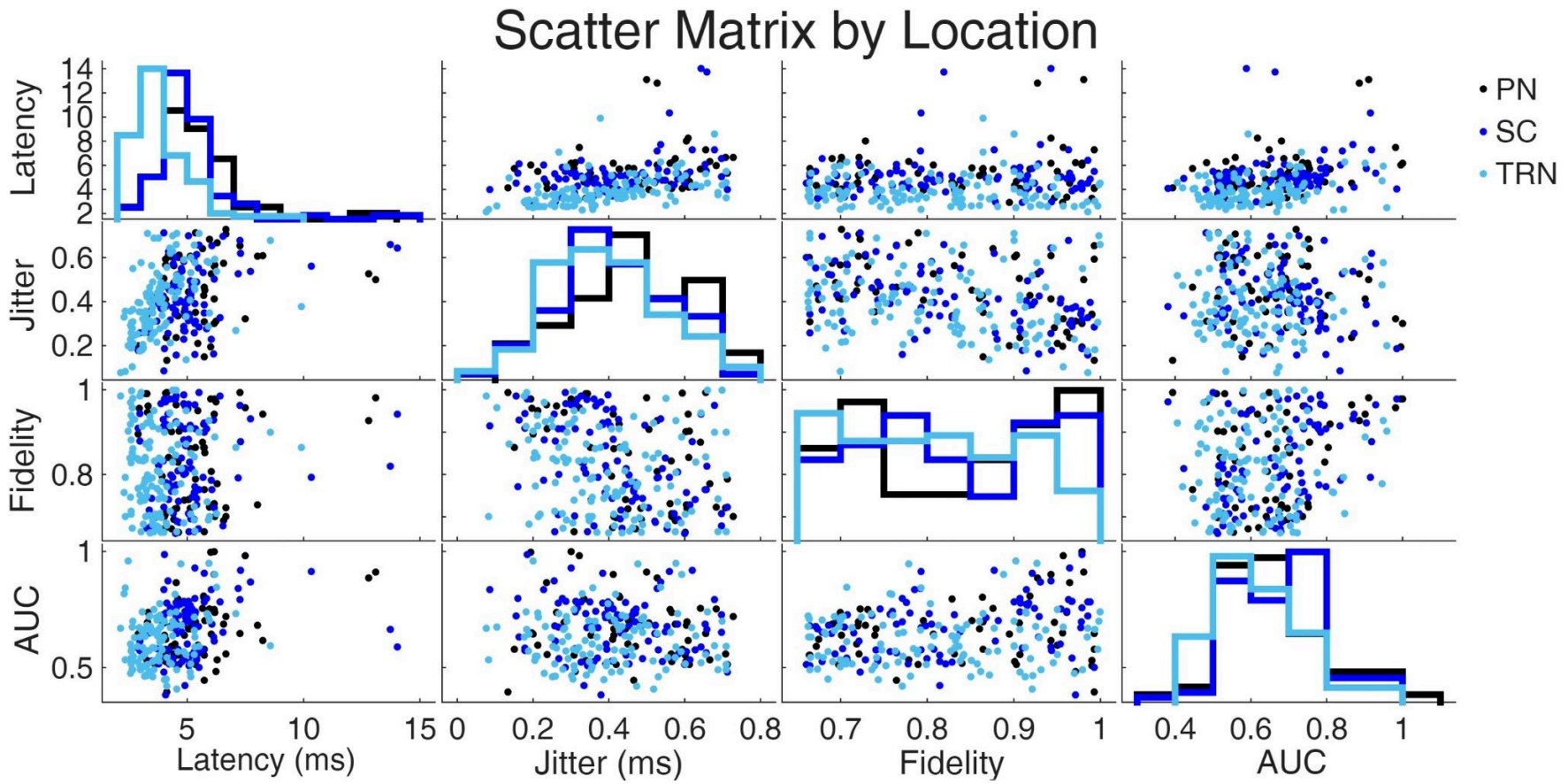

Figure 15 | Scatter Matrix of all Optotagged Units that Pass Criteria. Pairwise relationships between: latency, jitter, fidelity and AUC. OFC units are coloured by projection target region: PN (black), SC (dark blue),TRN (light blue). AUC = area under the curve. Note: for latency, jitter, fidelity and AUC thresholds See Methods 2.7.

The scatter matrix (Fig.15) had minimal separation between the three regions across all four metrics. This indicates that the optotagging method had good consistency and that the light-evoked physiological properties for these regions are conserved. The TRN demonstrated marginally lower latency; this is likely due to its closer anatomical distance to the OFC. All optotagged units met latency, jitter, fidelity, and AUC thresholds.

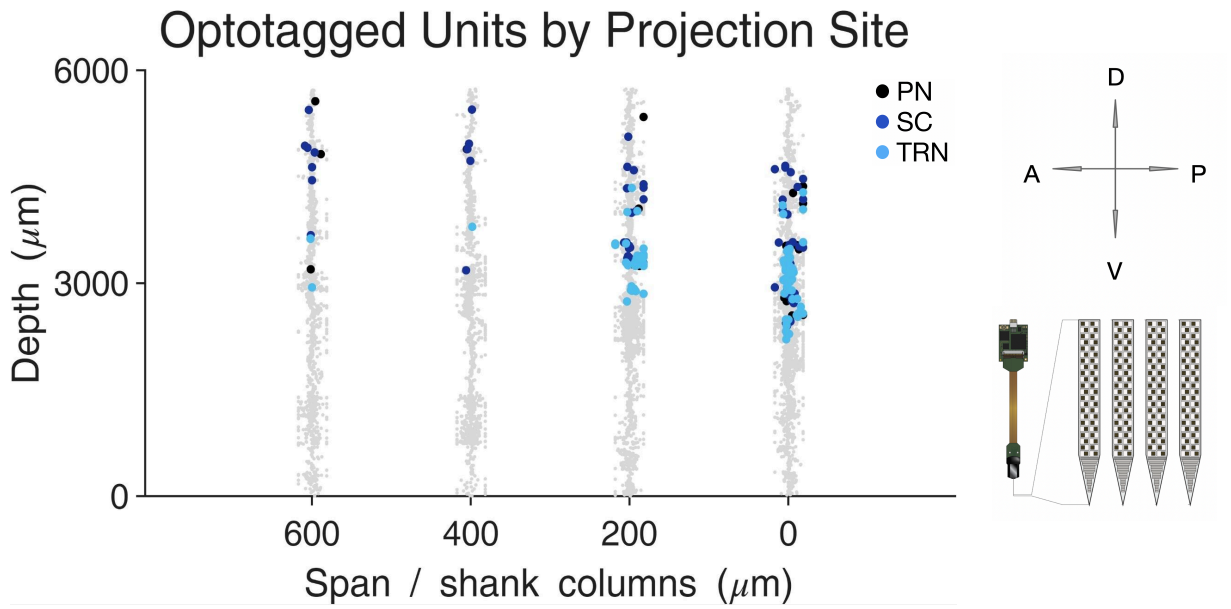

Figure 16 | OFC *Optotagged Units by Projection Site in Rat PMA104.* This figure shows all recorded units across the four electrode shanks. Depth is the length of the electrode from base (6000, dorsal) to tip (0, ventral), and Span marks the width between shanks in micrometers (600 being anterior with 0 being posterior). Status: grey dots = non-optotagged units, black dots = PN optotagged units, dark blue = SC optotagged units, and light blue = TRN optotagged units. On the right top corner is the positional axis that corresponds to the electrode layout where A = anterior, P = posterior, D = dorsal, and V = ventral. In the bottom right corner is an Electrode schematic for visual representation of the shanks.

Figure 16 illustrates that the majority of cells are situated in the dorsal and posterior regions of the electrode (upper end of shanks at span 200 and 0). TRN optotagged cells are predominantly clustered around the middle portions of the two shanks (200, 0), demonstrating spatial specificity. Based on prior knowledge that many cortical neurons from layer five project to the TRN (Caroll *et al.*, 2022), it is likely that this clustered arrangement for TRN-optotagged units is indeed in layer five. The PN and SC are also clustered in the same spatial locations but have more optotagged units in the anterior dorsal regions (See Fig.17 for individual spatial mapping). Given that there are no optotagged cells on the ventral portion of each of the four shanks, it is possible that the shanks are outside the OFC and instead inside the piriform cortex or olfactory bulb.

Optotagging Results Summary

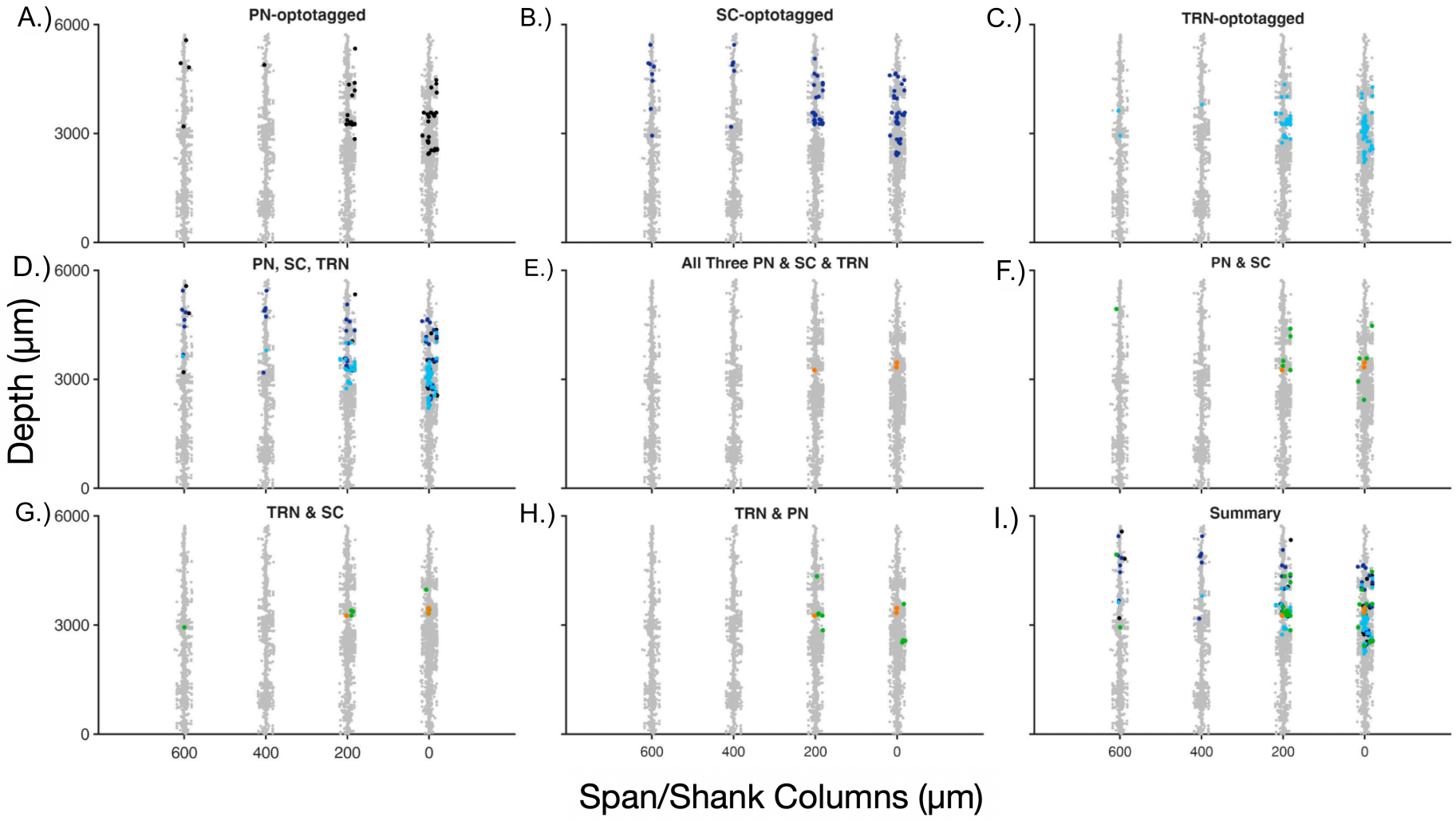

Figure 17 | *All Optotagged Units Across Shanks for Rat PMA104.* All recorded units are spatially positioned across the four electrode shanks. Depth is the length of the electrode from base (6000, dorsal) to tip (0, ventral), and Span marks the width between shanks in micrometers (600 being anterior with 0 being posterior). All non-optotagged units are colored in grey for each figure. A.) PN-optotagged units (black). B.) SC-optotagged units (dark blue). C.) TRN-optotagged units (light blue). D.) All units that are optotagged by the PN (black), SC (dark blue) and TRN (light blue). E.) Units that are optotagged by all regions in orange. F.) Units that are optotagged by PN & SC (green) and units by all three regions in orange. G.) Units that are optotagged by TRN & SC (green) and units by all three regions in orange. H.) Units that are optotagged by TRN & PN (green) and units by all three regions in orange. I.) Units that are optotagged by PN (black), SC (dark blue), TRN (light blue) and cells optotagged by at least two regions (green), and three regions (orange).

Figure 17 shows the spatial distributions of optotagged cells across shanks. Panels E, F, G & H illustrate that some optotagged cells are more responsive than others with projections to two or all three of the targets. The more responsive cells are typically situated on the dorsal and posterior portions of the shanks. Approximately 0.5% of the optotagged cells identified have projections to more than one region.

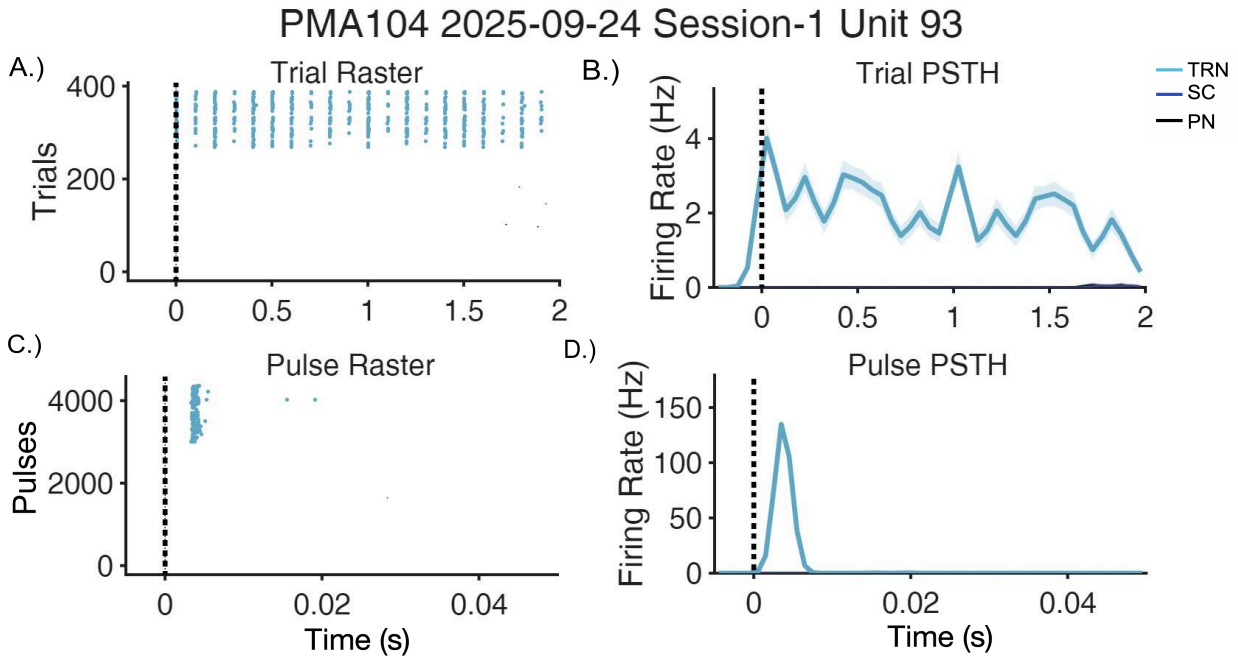

Figure 18 | *TRN Optotagged Unit 93 in Rat PMA104.* For panels A&B the dashed line is the onset of the trial, and panels C&D the dashed line is the laser pulse onset and each dot represents a spike. A.) The trial raster shows the spikes elicited during a trial over time (s). B.) Trial PSTH: the averaged and smoothed firing rate over time (s). C.) Pulse raster shows spike responses for each pulse D.) Pulse PTSH shows peak firing rate over time (s). *Note:* during each trial a train of 12 pulses with varying time, frequencies, and power are evoked (*See Methods* 2.7).

Figure 18 shows that this OFC unit 93 is responsive to light stimulation in the TRN, with spikes in the trial raster (Fig.18A) occurring in a vertical uniform pattern that is evidently time-locked to the light pulse trains. The pulse raster (Fig.18C) demonstrates good clustering after timepoint 0, meaning that the unit fired consistently with a rapid latency post light stimulation (< 0.01ms). This is supported in the Pulse PSTH (Fig.18D) where the peak firing also occurs at the same rapid timescales (< 0.01ms). This unit was completely unresponsive to the SC and PN, showing exclusivity to the TRN.

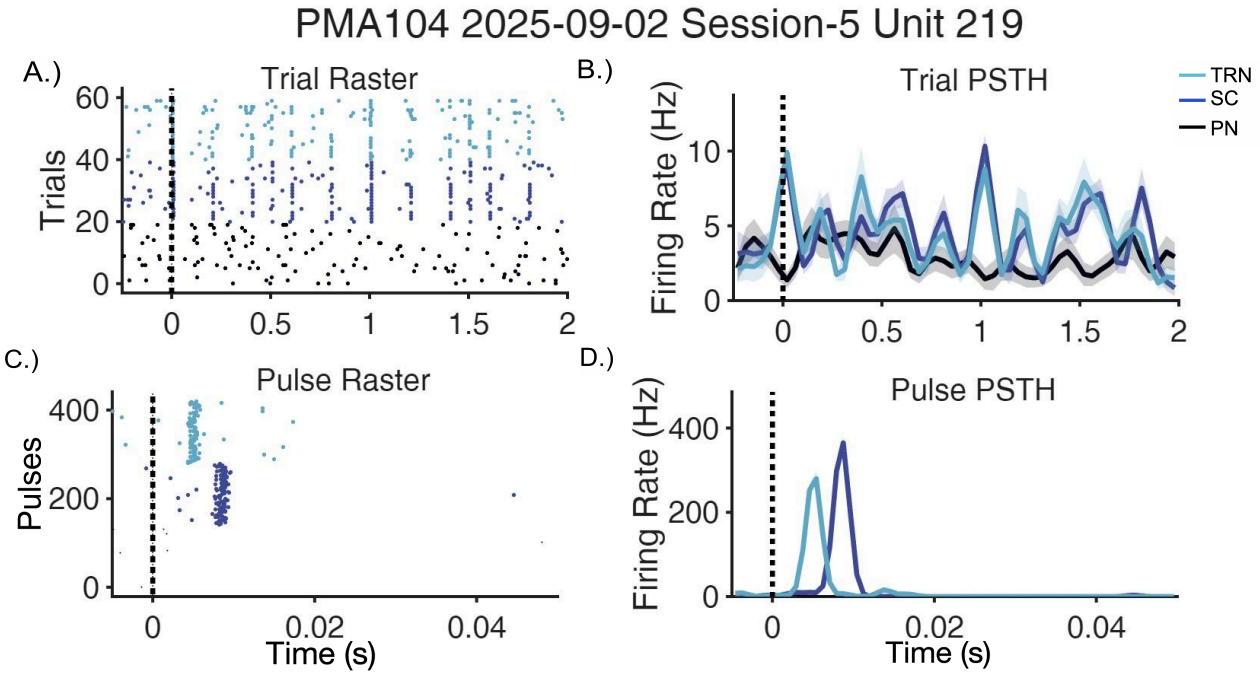

Figure 19 | *Thalamic Reticular Nucleus & Superior Colliculus Optotagged Unit 219 in Rat PMA104.* For panels A&B the dashed line is the onset of the trial, and panels C&D the dashed line is the laser pulse onset and each dot represents a spike. A.) The trial raster shows the spikes elicited during a trial over time (s). B.) Trial PSTH: the averaged and smoothed firing rate over time (s). C.) Pulse raster shows spike responses for each pulse D.) Pulse PTSH shows peak firing rate over time (s). *Note:* during each trial a train of 12 pulses with varying time, frequencies, and power are evoked (*See Methods* 2.7).

Figure 19 is a good example of a unit that was optotagged by two regions. In this case unit 219 was responsive to light stimulation in the TRN and SC but not the PN. This is demonstrated in tightly clustered spikes in the Pulse raster after time 0 (Fig.19C), and in the Pulse PSTH peaks, which show a sharp rise in firing rate post light onset for the TRN and SC conditions (Fig.19D), whereas this pattern is completely absent in the PN.

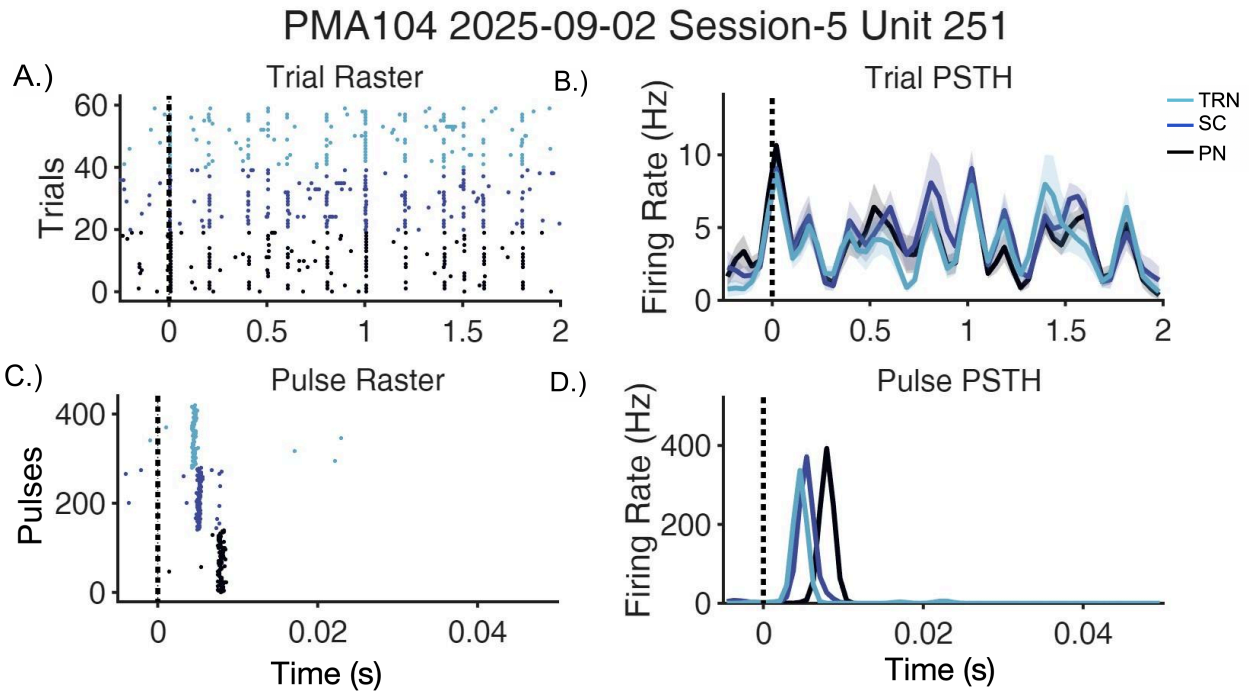

Figure 20 | *Unit 251 Optotagged by all Three Projection Sites in PMA104.* For panels A&B the dashed line is the onset of the trial, and panels C&D the dashed line is the laser pulse onset and each dot represents a spike. A.) The trial raster shows the spikes elicited during a trial over time (s). B.) Trial PSTH: the averaged and smoothed firing rate over time (s). C.) Pulse raster shows spike responses for each pulse D.) Pulse PTSH shows peak firing rate over time (s). *Note:* during each trial a train of 12 pulses with varying time, frequencies, and power are evoked (*See Methods* 2.7).

Figure 20 is a good example of a unit that was optotagged by all three regions. The trial raster has a clean response to each of the light trains; this is marked by the vertically aligned patterns (Fig.20A). The pulse raster and pulse PSTH (Fig.20 C&D) show that response latencies and peak firing rates occurred on rapid timescales for each of the three regions (<0.02s). There is a topologically evident response latency shown in the pulse raster, where the TRN was fastest followed closely by the SC, with the PN having the slowest response. This order is consistent with the anatomical distance of each region from the OFC.

Figures 18, 19, and 20 exemplify that responses to light are variable with some cells projecting to multiple areas. These examples are complementary to the data shown in figures 13 and 17 which give the full descriptive statistics of projection affinities and the mapping of their spatial locations respectively.

3.3 Ramp & Collapse Cells

Across 24 recording sessions in rat PMA104, a total of 19 R&C cells were identified: 13 with a ramp score of 1 and 6 with a ramp score of 2 (Fig.21). From this selection of R&C cells, three of these were optotagged. This section details the neuronal properties, spatial localisation, and projections sites for the three R&C optotagged cells.

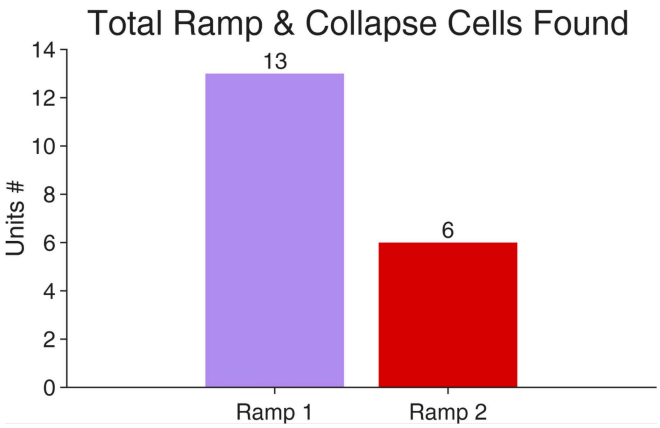

Figure 21 | Total Ramp and Collapse Cells Found in Rat PMA104. Ramp and collapse neurons with a ramp score of 1 shown in purple. Ramp and collapse cells with a ramp score of 2 show in red.

Note: See Methods for classification details

All cells mentioned in Figure 21 with the exception of one, were located centrally on the posterior shanks (Span: 200 and 0). This spatial layout is shown below in Figure 22. Out of the 19 identified R&C cells three were optotagged: units 91 and 298 had a ramp score of 2, with 102 having a ramp score of 1 (Fig.24). Figure 23 shows unit 39 as an excellent example of a ramp and collapse profile found during analysis of PMA104.

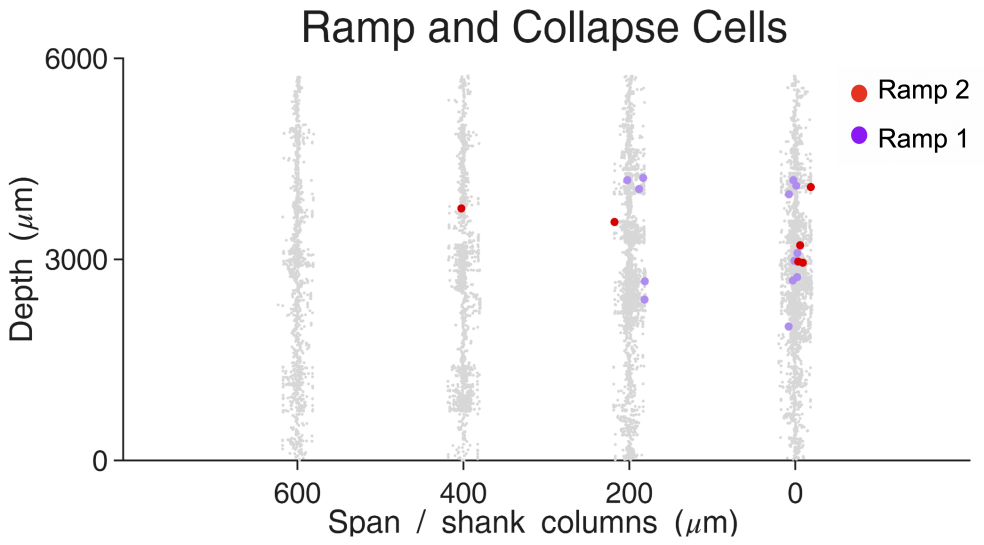

Figure 22 | Ramp and Collapse Cell Locations for Rat PMA104. All R&C recorded units across the four electrode shanks. Depth is the length of the electrode from base (6000, dorsal) to tip (0, ventral), and Span marks the width between shanks in micrometers (600 being anterior with 0 being posterior). Status: grey dots = non-optotagged units; red dots = ramp score 2 units, purple dots = ramp score 1 units.

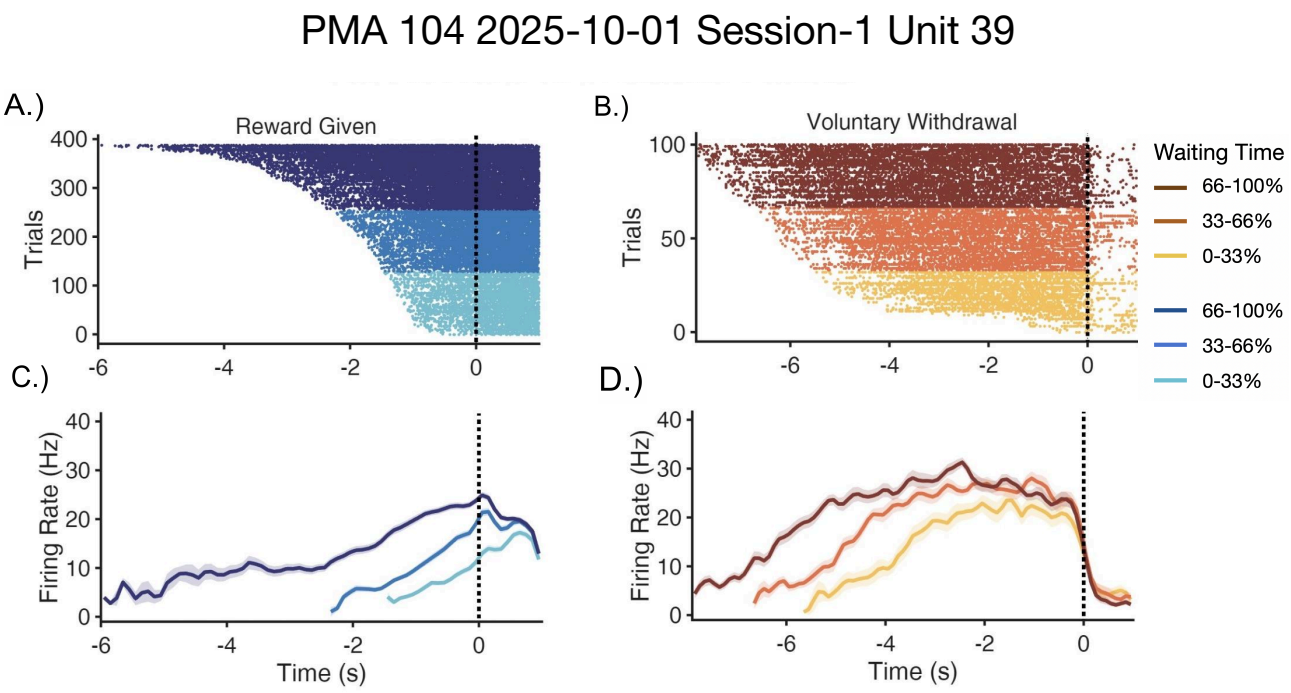

Figure 23 | Unit 39: Firing Patterns of Ramp and Collapse Cell During Abandoned and Rewarded Trials in Rat PMA104. PSTH (100-ms bins, mean ± SEM) and raster plots from a ramp/collapse cell recorded with a neuropixel 2.0 silicon probe. Activity aligned to trial end (abandonment and reward).

The 19 R&C cells identified in this study exhibited consistent ramping and collapsing firing trajectories. Presented here in Figure 23 is one of these examples. Unit 39 had ramping that occurred on longer timescales (>5 seconds), ramping was steep with a firing rate peak at ~30Hz followed by an abrupt collapse before the decision to leave the port at time 0 (bottom right figure). Rewarded trials mirrored this ramping pattern until reward interruption occurred. This data is in support of preliminary findings by Lagler *et al., (n.d).*

3.4 Optotagged Ramp & Collapse Cells

As aforementioned, out of the 19 R&C cells identified, three of these were optotagged. In this section both the ramp and collapse profiles and the optotagging results are presented for the optotagged units 91, 298 and 102 respectively.

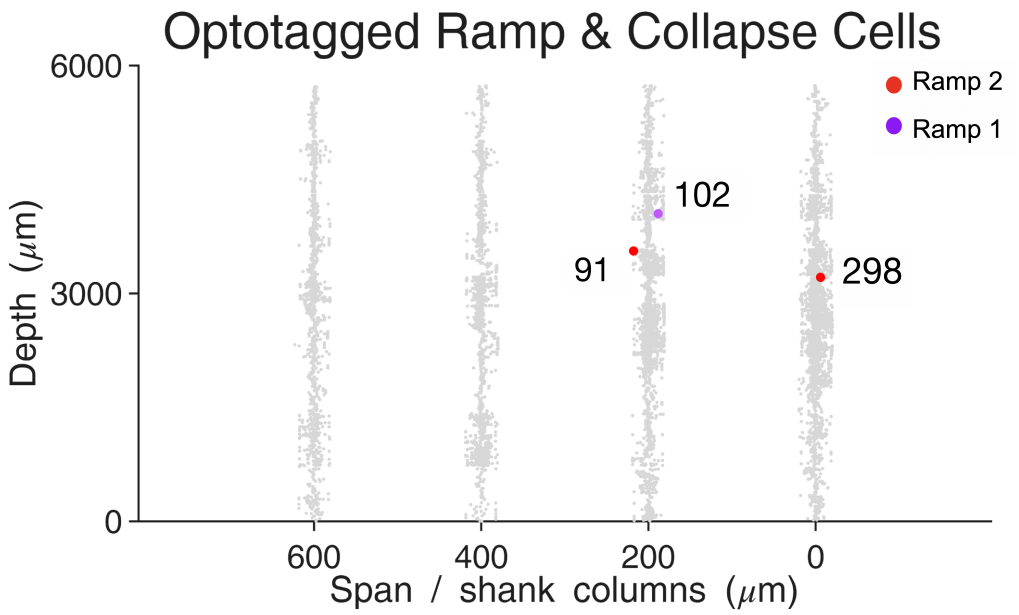

Figure 24 | Optotagged Ramp and Collapse Cell Locations for Rat PMA104. All R&C cell optotagged units across the four electrode shanks. Depth is the length of the electrode from base (6000, dorsal) to tip (0, ventral), and Span marks the width between shanks in micrometers (600 being anterior with 0 being posterior). Status: grey dots = non-optotagged units; red dots = ramp score 2 units, purple dots = ramp score 1 units. Note: Ramp cells are enlarged for better visibility compared to all units.

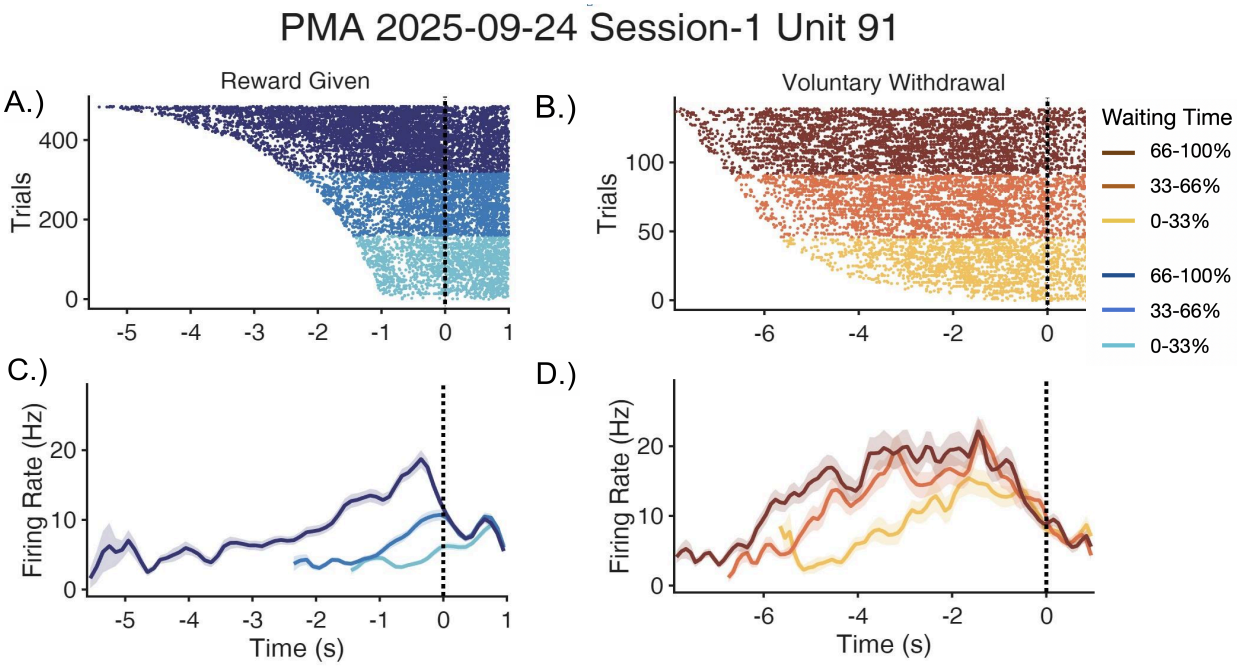

Figure 25 | Unit 91: ***Firing Patterns of Optotagged Ramp and Collapse Cell During Abandoned and Rewarded Trial in Rat PMA104.*** This cell is an optotagged cell with a ramp score of 2. PSTH

(100-ms bins, mean ± SEM) and raster plots from a ramp/collapse cell recorded with a neuropixel 2.0 silicon probe. Activity aligned to trial end (abandonment and reward).

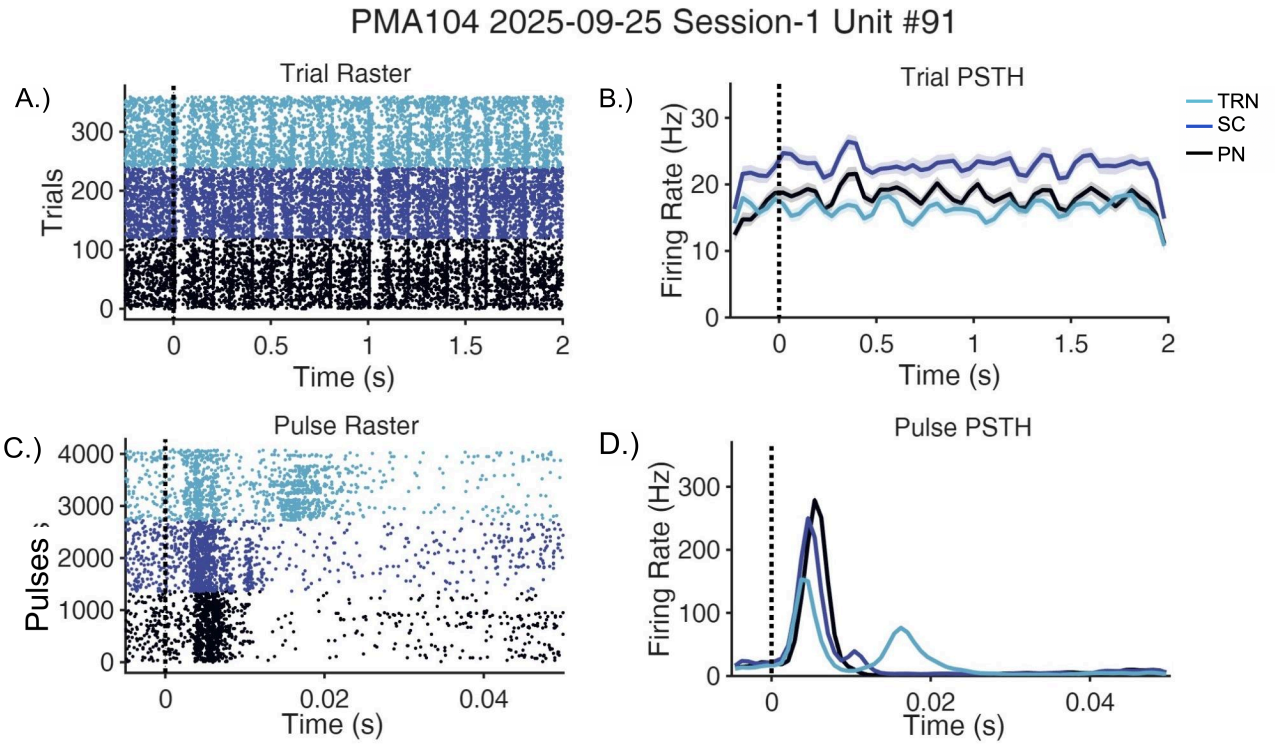

Figure 26 | Ramp & Collapse *Unit 91: Optotagged by all Three Projection Sites in PMA104.* For panels A&B the dashed line is the onset of the trial, and panels C&D the dashed line is the laser pulse onset and each dot represents a spike. A.) The trial raster shows the spikes elicited during a trial over time (s). B.) Trial PSTH: the averaged and smoothed firing rate over time (s). C.) Pulse raster shows spike responses for each pulse D.) Pulse PTSH shows peak firing rate over time (s). *Note:* during each trial a train of 12 pulses with varying time, frequencies, and power are evoked (*See Methods* 2.7).

The R&C profiles of unit 91 (Fig.25) convey gradual ramping over time with peaks between ~10-25Hz and a collapse to ~8Hz in abandoned trials. In reward conditions there are evident ramping slopes, however, on the longer wait times (Fig.25C) there is a moderate collapse that gets interrupted with reward. It is likely that the animal was considering abandoning the task around this time. The ramping and collapsing profiles align with preliminary R&C reports by Lagler *et al.* (*n.d.*). Unit 91 demonstrated a strong activation in response to light stimulation in all three projection targets (PN, SC & TRN). This was marked by rapid firing responses post light onset (< 0.01ms) in the pulse raster and pulse PSTH (Fig.26). The bulk of trial spikes in the pulse raster were clustered together, with some jitter, chiefly in the TRN (Fig.26C). It is feasible that the clear separation of firing clusters for the TRN is evidence of monosynaptic-induced firing. Spikes occurred in a vertically uniform pattern that was in alignment with the 12 light pulse trains (Fig.26A).

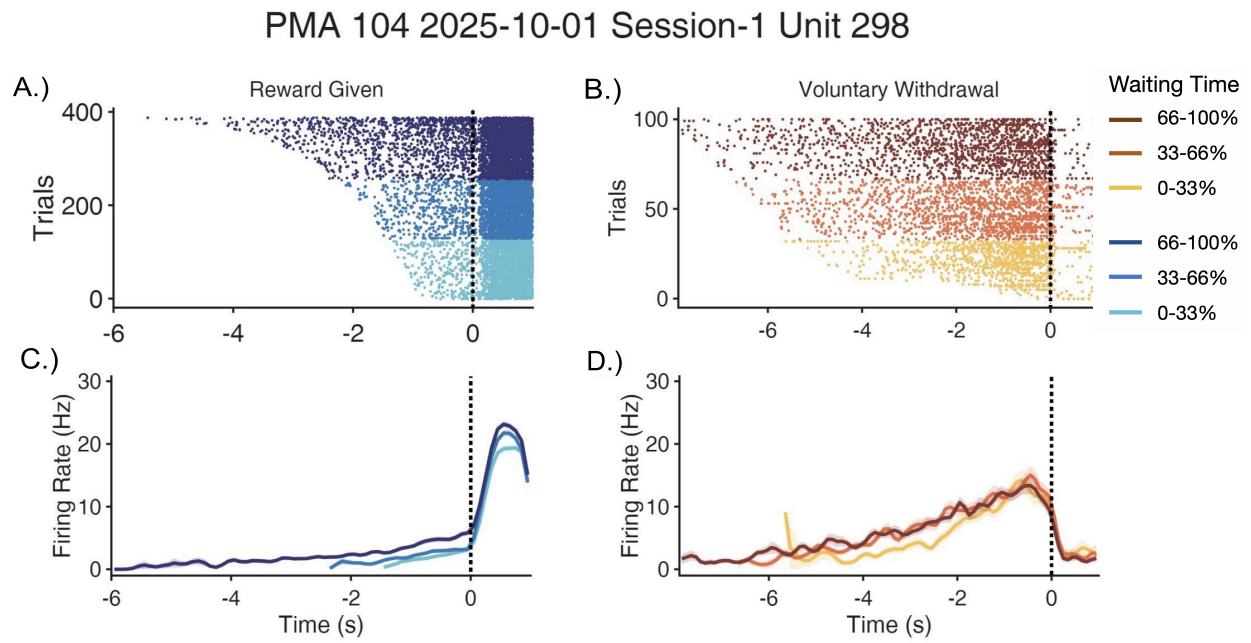

Figure 27 | Unit 298: ***Firing Patterns of Optotagged Ramp and Collapse Cell During Abandoned and Rewarded Trial in Rat PMA104.*** This cell is a TRN optotagged cell with a ramp score of 2. PSTH (100-ms bins, mean ± SEM) and raster plots from a ramp/collapse cell recorded with a neuropixel 2.0 silicon probe. Activity aligned to trial end (abandonment and reward).

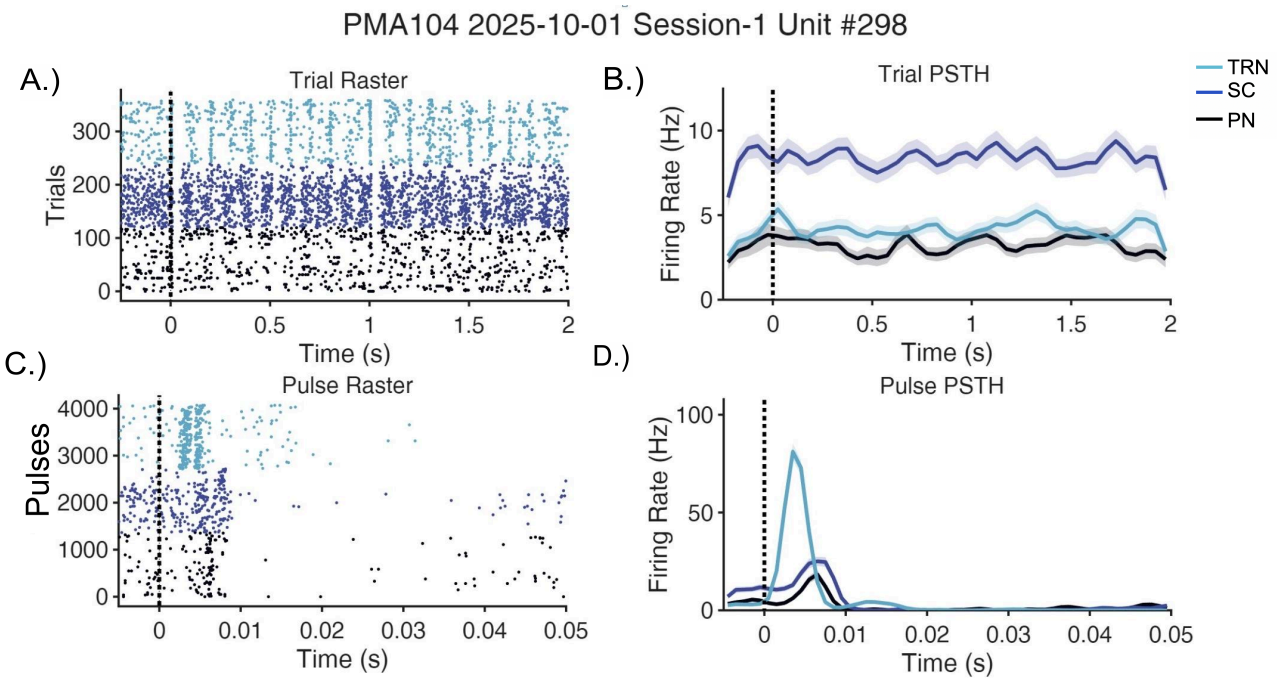

Figure 28 | Ramp & Collapse *Unit 298: Optotagged by the Thalamic Reticular Nucleus in PMA104.* For panels A&B the dashed line is the onset of the trial, and panels C&D the dashed line is the laser pulse onset and each dot represents a spike. A.) The trial raster shows the spikes elicited during a trial over time (s). B.) Trial PSTH: the averaged and smoothed firing rate over time (s). C.) Pulse raster shows spike responses for each pulse D.) Pulse PTSH shows peak firing rate over time (s). *Note:* during each trial a train of 12 pulses with varying time, frequencies, and power are evoked (*See Methods* 2.7).

R&C unit 298 (Fig.27) interestingly, seems to be reward modulated due to the pronounced rise in firing rate (up to ~20Hz) in the second of reward delivery (Fig.27C). Unit 298 has a uniform time-locked trial pattern for TRN trains (Fig.28A). The pulse raster (Fig.28C) shows that the TRN spike activity post-light stimulation is clustered shortly after time 0, with corresponding firing rate increases during this epoch reflected in the pulse PSTH (Fig.28D). There is some evidence that unit 298 is optotagged by the PN and SC as well; noticeable clusters post light delivery occur <1 ms in time with parallel rises in firing rate (Fig.28, C&D). Nonetheless, this evidence for PN & SC is not sufficient to pass the optotagged criteria.

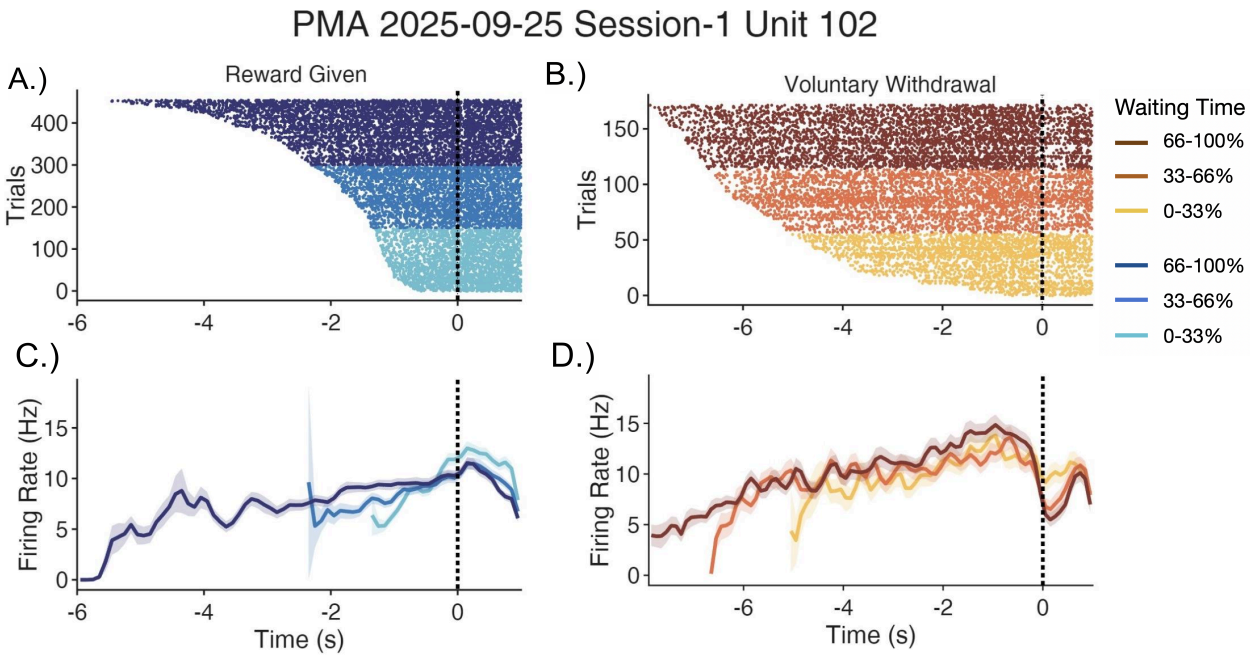

Figure 29 | Unit 102: *Firing Patterns of Optotagged Ramp and Collapse Cell During Abandoned and Rewarded Trial in Rat PMA104.* This cell is a PN & SC optotagged cell with a ramp score of 1. PSTH (100-ms bins, mean ± SEM) and raster plots from a ramp/collapse cell recorded with a neuropixel 2.0 silicon probe. Activity aligned to trial end (abandonment and reward).

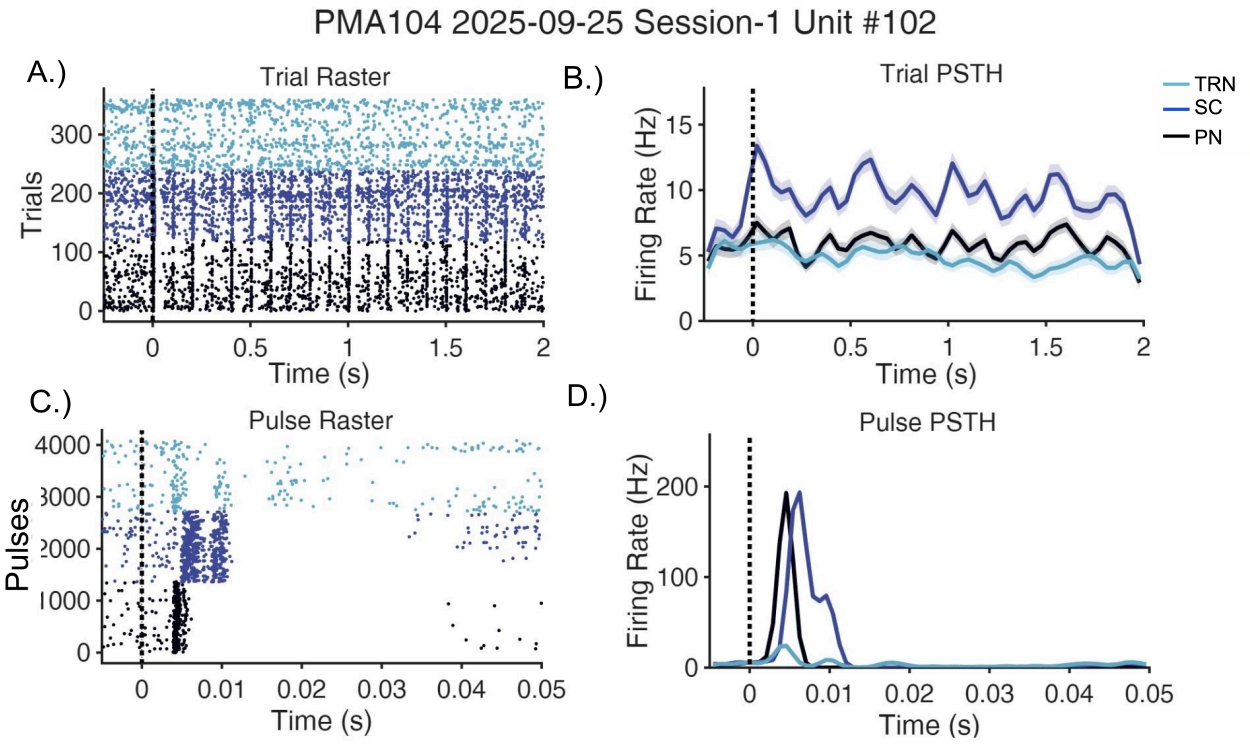

Figure 30 | Ramp & Collapse *Unit 102: Optotagged by the Pontine Nucleus and Superior Colliculus in PMA104.* For panels A&B the dashed line is the onset of the trial, and panels C&D the dashed line is the laser pulse onset and each dot represents a spike. A.) The trial raster shows the spikes elicited during a trial over time (s). B.) Trial PSTH: the averaged and smoothed firing rate over time (s). C.) Pulse raster shows spike responses for each pulse D.) Pulse PTSH shows peak firing rate over time (s). *Note:* during each trial a train of 12 pulses with varying time, frequencies, and power are evoked (*See Methods* 2.7) .

The Ramp unit 102 presented in figure 29 had a ramp score of 1. The ramp score of 1 was due to the less exaggerated collapse evident in the voluntary withdrawal condition (Fig.29D), post abandonment (after time 0) the trajectory showed a rise in firing activity. As shown in figure 30, unit 102 had successful optotagging for the PN and SC; this was marked by well-clustered spikes and a sharp increase of peak firing rates post light onset in the pulser raster and pulse PSTH respectively (Fig.30, C &D). The TRN was not optotagged for this unit.

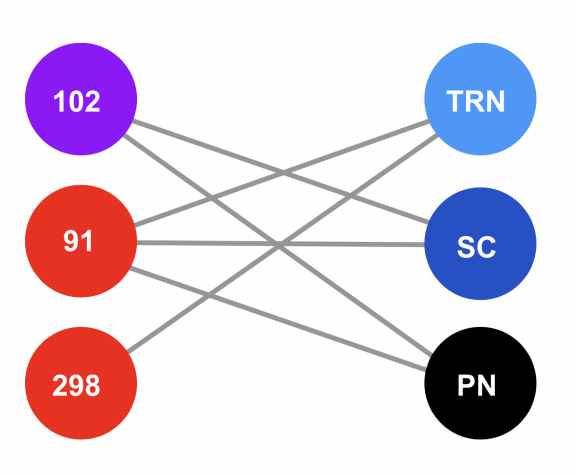

Figure 31 | Network Summary for Optotagged Ramp & Collapse Cells. Visual representation of the network results discovered in this thesis. Circles = nodes, grey lines = edges. Each line represents the confirmed projection from unit to brain target region. Optotagged R&C unit 102 with a ramp score of 1 shown in purple. Optotagged R&C units 91 & 298 with a ramp score of 2 shown in red. Projection area TRN shown in light blue. Projection area SC shown in dark blue. Projection area PN shown in black.

3.5 Results Summary

Vevaiometric, wait time distributions, and psychometric plots demonstrated that the animals possessed confidence-dependent waiting behaviour during the 2AFC task (3.0 Behaviour). This aligned with previous research findings (Kepecs *et al.*, 2008; Lak *et al.*, 2014; Masset *et al.*, 2020). Histology showed that animals had sufficient viral expression, and implants in the correct locations; however, only animal PMA104 had successful optotagging (Table 1). Across all of the optotagging trials for PMA104, a total of 184 purely optotagged units were found: with 47 in the PN, 76 in the SC, and 96 in the TRN (3.2 Optotagging). A fraction of these optotagged cells had projections to two or all three regions. The majority of optotagged cells clustered on the posterior portions of the electrode shanks, which likely corresponds to layer 5 of the cortex. The TRN optotagged cells had slightly faster response latencies; however, there was minimal difference between the three regions across jitter, fidelity, and AUC measures. Thus, the optotagging method was consistent. From the electrophysiology recordings obtained during the 2AFC task in animal PMA104, 19 R&C cells were identified (3.4 Optotagged Ramp and Collapse Cells). Of these cells, three of them were optotagged: units 91, 298 and 102. The R&C optotagged unit 91 was responsive to all three regions. The R&C optotagged unit 298 was exclusively responsive to the TRN and seemed to be reward modulated. Finally, the R&C optotagged unit 102 was responsive to the PN and SC but not the TRN. These connections are summarised in figure 31. The Optotagged R&C cells were spatially located on the middle portions of the posterior shanks.

In sum:

I found OFC Ramp and Collapse cells that project to the PN, SC and TRN (Fig.31). These cells were predominantly located in the central/posterior portions of the electrode shanks (Fig.24). This organisation is congruent with the hypothesis that functionally defined OFC neurons, including Ramp & Collapse cells, will show a spatial and anatomical specificity in their projection targets.

Discussion 4

In this thesis I have provided evidence that the rats exhibit confidence-dependent waiting behaviour, this is in alignment with previous findings (Kepecs *et al.*, 2008; Lak *et al.*, 2014; Masset *et al.*, 2020). In rat PMA104 I have confirmed OFC projections to three brain regions: the PN, SC and TRN via the antidromic optotagging method. For rat PMA104, a total of 19 R&C cells were detected; this is in support of preliminary findings by Lagler *et al.* (*n.d.*) who first identified R&C cell phenomena. Of the R&C cells identified, three of them were optotagged, two were responsive to multiple target regions and one to the TRN only. For the first time I have confirmed that the novel R&C cells project to the PN, SC and TRN.

4.1 The Story of Ramp and Collapse Cells

19 R&C cells were found. This further supports that the preliminary findings by Lagler *et al.* (*n.d.*) are reproducible and validate the existence of R&C cell phenomena. Only 1-8% of recording units are estimated to be R&C cells; the chance that these cells are also optotagged near the electrode is low, making finding such cells a rarity. I managed to find three optotagged R&C cells. I am well aware that the plural of anecdote is not data, and that this work leads to more questions than it gives answers. However, I do have three interesting anecdotes to share, which I trust will lay the groundwork to be added upon with future scientific work.

First, I will start with the optotagged ramp and collapse unit 91: this unit had a ramp score of 2 and was optotagged by light in all three target regions: PN, SC, and TRN (Figs.25 & 26). From my optotagging data, only three cells (comprising a mere 0.04% of all units recorded) were responsive to all three regions, and unit 91 is one of them. Based on this finding, one can infer that some R&C cells are possibly rich club neurons: highly connected integrative hubs that facilitate communication across multiple brain regions (Van den Heuvel & Sporn, 2011). This is fitting given that rich club neurons are often active during higher-order cognitive tasks such as decision-making, task switching, valuation and adaptive behaviour (Grayson *et al.*, 2014; Mecklenbrauck *et al.*, 2023; Senden *et al.*, 2017; Van den Heuvel & Sporn, 2011). Future studies should aim to determine if R&C cells are highly connected to other brain regions to confirm rich club properties. If R&C cells are indeed rich club neurons, their function may extend further than confidence encoding, warranting investigation of their activity in multiple contexts. The TRN pulse raster and PSTH (Fig.26, C & D) for Unit 91 had a second wave of activity at approximately 0.02 s post light onset. This delayed spiking may arise from looped excitation between TRN and thalamic relay neurons, implying that this unit contributes to feedback-driven network activity rather than just feedforward transmission. In essence, unit 91 was highly connected and possibly involved in a recurrent network.

Second, is the story of the ramp and collapse optotagged unit 298 (Figs.27 & 28): this cell had a ramp score of 2 and was purely optotagged by the TRN only. However, there is some response sensitivity to both the PN and SC (Fig.28, C & D) where the cell shows clustered activity <0.1s in the pulse raster and moderate elevation of firing in the pulse PSTH for these areas. Nonetheless, this is not sufficient to pass the optotagging criteria. A potential reason for why responses from the PN and SC were weak could be that the light delivery to the target was compromised or that viral expression in the OFC neuron was low. What is fascinating about unit 298 is that it shows a clear reward-modulated response, as demonstrated by a sharp rise in firing rate at the moment of reward delivery. Thus, 298 could potentially encode reward outcomes. A valuable future analysis would be to determine if other ramp cells also exhibit reward-related activity. As more R&C cell data is collected, it is likely that greater heterogeneity across ramping cells will emerge. In sum, unit 298 is a TRN-optotagged reward-responsive neuron, and it is possible that optotagging of the PN and SC for this unit failed due to mechanical reasons.

Third, unit 102 (Figs.29 & 30) had a ramp score of 1 and was purely optotagged by two regions: the PN and the SC. This neuron had a less dramatic collapse in the voluntary withdrawal condition, and this mild collapse was followed with a delayed increase in firing after time 0 (Fig.29D). Could it be that this cell is encoding post-decision outcomes, reward vs no reward, similar to unit 298? Or perhaps the animal changed its mind; after initially abandoning the task, it decided to re-engage by reentering the port to wait for reward delivery. This possibility would need to be confirmed with the video footage collected for that trial. Evidence for optotagging in the TRN was weak, but not absent. Akin to unit 298, it is possible that the optotagging failed due to mechanical reasons. Unit 102 also had a second wave of spiking activity that was exclusive to the SC; this activity was under 0.01s, which means that this activity was probably jitter as opposed to circuit-looped activation. Unit 102 had less characteristic ramping profiles compared to 91 and 298 which were class 2 ramp cells. Something to ponder on in regard to this is why do some R&C cells have less distinguished ramping and collapsing trajectories. Is it due to a decreased level of confidence? Or is it simply explained by differing physiological conditions such as varying network inputs and intrinsic membrane properties?

The three optotagged units 91, 298, and 102 had differing firing patterns and projection targets. Unit 91 was responsive to the light stimulus in all target regions, unit 298 was only responsive to the TRN and had reward-modulated firing, and unit 102 was responsive to the PN and SC with a less distinct ramping and collapsing trajectory. These differences evince that R&C cells have heterogeneous ramping dynamics. Thus, these three anecdotes open up the question of why there is such diversity and what does this mean in the context of confidence computation?

4.2 Decoding Ramp and Collapse Activity

Ramping activity, a gradual increase in firing activity over time, is a widespread firing pattern observed across many brain regions during behavioural paradigms, including decision-making and memory (Brody, 2003), during motor preparation (Churchland, Santhanam & Shenoy, 2006;

Kaufman, Churchland & Shenoy, 2014) and in the perception of time (Jazayeri & Shadlen, 2015; Merchant & Averbeck, 2017). This ramping firing pattern is thought to be integrating information over time. In motor regions, ramping neurons often reach their peak firing rate at the onset of movement; in this context, the signal is interpreted to be involved in motor planning and the execution of movement (Churchland, Santhanam & Shenoy, 2006). To clarify, R&C cells are not motor related; R&C cells, in contrast, have a distinct pattern: from the onset of waiting, their ramping activity builds up over time and abruptly collapses *prior* to task abandonment/movement onset. Thus, the timing of collapse is unique to R&C cells. Rather than driving movement execution, these cells are encoding internal signals such as confidence and temporal tracking that precede action.

Another way to think of R&C activity is through the lens of frustration or motivational disengagement, as opposed to confidence. The ramping phase could reflect a gradual increase in anticipatory drive during reward expectation. Since rats in this task are water deprived, waiting even a few seconds, let alone up to eight seconds, is likely to induce increasing frustration over time, which may be reflected in this ramping activity. The eventual collapse in firing could be a result of motivational fatigue, when the rat gives up on the task and decides to start a new trial. In other words, R&C cells may be cost-benefit analysing cells, which track temporal investment (cost) and evaluate reward outcome (benefit). When the benefit doesn't materialise, then collapse; a signal to downstream neurons to abandon the task. This interpretation would be in alignment with prior research establishing that the OFC is involved in reward magnitude, effort, time, and valuation (Fatahi *et al.*, 2018; Fellows, 2011; Schnitzler & Wojtecki, 2012). Further supporting this theory is that unit 298 as described above, was reward modulated suggesting that some ramp cells may encode reward outcomes. What this interpretation of R&C cells does not account for is confidence. How does R&C cell activity change in response to sensory evidence?

In the preliminary work by Lagler *et al.* (*n.d.*), the R&C ramp slope scales with waiting duration, where steeper ramps predict shorter waits, whilst shallower ramps predict longer waits. The neural activity patterns of R&C cells reflect confidence over time; this is supported in psychometric, vevaiometric, and calibration curves by Lagler *et al.* (*n.d.*). Laglers' work is consistent with Bayesian models, which define confidence as the probability of being correct given available evidence (Meyniel *et al.*, 2015). As aforementioned there was heterogeneity across my three optotagged ramping cells. To understand what is happening here we need more data, coupled with a deeper analysis to extract a more meaningful and evidence-based interpretation of R&C confidence encoding properties within a network. Future analyses should explore different brain areas that R&C cells project to. Once sufficient data is acquired, to find order within the vastness of high dimensional multi-neuron datasets one should adopt dimensionality reduction techniques for this analysis.

Dimensionality reduction techniques are mathematical methods that compress high-dimensional neural recordings into a small set of latent variables that preserve the original patterns and dynamics of the data (Ray, Reddy & Banerjee, 2021). The power of the technique is that it allows one to see population dynamics that are not seen in single-neuron activity alone, as

well as to match these dynamics to behaviour. One such technique is Latent Factor Analysis via Dynamical Systems (LFADS); a non-linear artificial non-recurrent neural network (RNN) lauded for its ability to account for single-trial phenomena such as behavioural variability and spike fluctuations in single neurons (Pandarinath *et al.,* 2018). Boucher *et al.* (2023) used Latent Factor Analysis via Dynamical Systems (LFADS) to decode and predict behavioural metrics during a checkerboard task in monkeys, finding that the technique improved the prediction of reaction times and choice by decoding initial conditions in the premotor cortex. The results demonstrated that neural trajectories are modulated by sensory evidence and initial brain states. The LFADS system, due to its robust applicability to premotor signals has shown potential for use in brain-machine interfaces and robotics, underscoring its merit in decoding large sets of neural data (Botadra *et al.,* 2022; Li *et al.,* 2020; Ran *et al.,* 2022). Bussell *et al.* (2023) observed OFC neuronal population activity responsive to information and reward in mice during an odour task using a dimensionality reduction technique termed CEBRA (Schneider *et al.*, 2023). Low-dimensional dynamics that emerged with the odour task behaviour advocated that mice have different neural pathways representing the intrinsic value of information and the intrinsic value of reward, and that mice like to acquire knowledge, preferring it over reward in some instances. With the OFC encoding the quality of such knowledge acquired.

Much of the research using dimensionality reduction techniques has centred around premotor activity and few publishings exist on population dynamics underlying cognition and decision-making, making this direction of exploration an exciting pursuit. Applying these techniques to R&C analysis could help to explain the heterogeneity; what hidden dynamics are shared across R&C cells and what dynamics seem to be specialised and conserved within specific subpopulations. Latent embeddings could shed light on how R&C neural population states evolve over time, revealing transitions between internal processes such as evidence accumulation, confidence estimation, and outcome evaluation. With these techniques we may be able to answer questions like how do R&C neurons locally encode decision confidence, track temporal investment, and signal choice abandonment by integrating these variables across time? Whilst these techniques could be telling about R&C population dynamics they do not explain how these patterns occur across a global network.

Li and Wang (2021) showed that confidence can emerge from adaptive, network-level computations without relying on specialised neurons. Their model reproduces behavioural patterns similar to those in the 2AFC task, such as increased temporal investment with stronger evidence. The work outlined by the authors supports the notion that OFC confidence signals reflect distributed population dynamics. Building on this idea, if R&C cells turn out to be rich club neurons, it is possible that they are drivers/regulators of such population dynamics. To better understand the role of R&C cells, it is important to know whether they are necessary or sufficient for confidence-dependent waiting. If R&C cells were inhibited during the 2AFC task, would we see a decrease in confidence-dependent waiting time? Conversely, if the R&C cells were activated during waiting, would this result in longer wait times, especially during trials when evidence is low?

To address such questions one could use the newly developed Neuropixel-Opto probes, electrodes that enable simultaneous optical control of neurons *in vivo* (Lakunina *et al.*, 2025). By implanting multiple opto-probes in different regions, it would be possible to activate or inhibit a unit in one brain region (A) whilst recording the response in another region (B), as well as the reverse B → A, to confirm bidirectional monosynaptic connectivity and function. This type of manipulation could help determine what type of causal influence R&C cells have in both local and global networks and directly align this to behaviour. Another way to understand local network computations is through the use of monosynaptic inference. Techniques such as GLMCC (Kobayashi *et al.*, 2019) and CoNNECT (Kobayashi *et al.*, 2020) allow one to statistically identify putative monosynaptic connections from parallel spike trains. Applying such methods to R&C analysis could allow one to distinguish the local microarchitecture of R&C cells in the OFC and then relate this knowledge to global brain connectivity defined via the antidromic optotagging method or by using multiple neuropixel-opto probes.

R&C cells appear to be part of a distributed network involving the PN, SC, and TRN; what is unknown is how this network helps to integrate confidence, reward, and time, as well as what other regions are involved. Further, what has not been explored is if R&C cells are connected to other R&C cells. The 19 R&C cells found in this study were in close proximity to each other on the electrode shanks (Fig.22), though this does not necessarily mean they are connected, research has illustrated that neurons are more likely to be synaptically connected if they share similar functional properties (Ko *et al.*, 2011; Perin, Berger & Markram, 2011; Song *et al.*, 2005). Future academics should investigate how R&C cells as a population communicate with each other to encode confidence. By combining dimensionality reduction techniques with circuit-mapping tools, future research may be able to better understand how R&C cells compute confidence and process information across a network, as well as how this guides confident behaviour.

4.3 The Superior Colliculus

The SC was previously thought of as a rather simple structure that controls eye movements (Dorris, Paré, & Munoz, 1997). However, more recent understanding has established that the SC plays a larger role in sensory processing, spatial orientation, attention, decision-making and saliency computation (Basso & May, 2017; Cui *et al.*, 2025; Maunsell, 2015). The SC is known to communicate with cortical regions *indirectly* via the thalamus and basal ganglia (Cavanaugh, McAlonan & Wurtz, 2020; Redgrave, 2010). In this thesis, I have presented evidence that the OFC has direct connections to the SC (Fig.17B). This finding is supported by preliminary viral tracing studies conducted by Mirjam Lambourne who identified strong connections to the SC (fig.1). Building on this knowledge, I have identified two R&C cells in the OFC that project to the SC (figs.26 & 30). The key question here is what information is being encoded between these cells?

It is unclear how Bayesian-like computations are distributed across neural circuits; however, authors Basso & May (2017) postulated that the superior colliculus may play a role in such behaviour. Without further investigation I can only give speculative answers. Perhaps the connections between the OFC (including R&C cells) and the SC help to encode saliency and attention cues relevant to confidence-dependent tasks. It is feasible that the strength of SC signals (encoding higher saliency and attention) is modulated by OFC cells. If so, this could impact the time an animal is willing to invest in waiting for a reward. This speculation would need to be investigated via inhibition of this circuit during the 2AFC task to see whether waiting times or choice accuracy are impacted by the disruption. This logic comes from a study by Jun *et al*. (2021) who demonstrated that when the SC was inactivated in primates via microinjections of lidocaine or muscinol, this led to biased choices and decreased evidence processing in a perceptual decision-making task. Adjacent to this, a study by Thomas *et al.*, (2023) found that the SC modulates choice competition and drives choice activity in the frontal cortex. Therefore it is plausible that SC inhibition during the 2AFC task in rats could lead to changes in choice accuracy and waiting behaviour; proxies for confidence.

The SC is involved in sensorimotor transformation (Massot, Jagadisan & Gandhi, 2019). Notably, the SC contains a topographic map of auditory space which helps localise where a sound is coming from and guides body orientation towards or away from it (Palmer & King, 1982). Movements like repositioning the head and body towards salient auditory stimuli, for example towards the port associated with the higher click trains, may depend on SC encoded localisation and saliency, as well as input from the OFC to determine if these responses are appropriate. Research centered on the SC predominantly establishes attention and saliency for visual processing. An intriguing future investigation would be to define the OFC projections to the inferior colliculus via optotagging, for this brain area is better adjusted to auditory processing, and thus would be arguably more relevant to the 2AFC task.

4.4 The Thalamic Reticular Nucleus

The TRN is a fundamental node connecting the thalamus and cortex, playing a major role in synchronising corticothalamic rhythms, modulating attention, and the processing of sensory information (Campbell, Govindaiah & Guido, 2024). The TRN is a cluster of GABAergic neurons that provides inhibitory feedback within the thalamocortical loop, allowing it to shift neuronal firing patterns from tonic single spikes to low-threshold bursting (Willis *et al*., 2015). This transformation modulates the sensitivity of thalamocortical neurons, controlling whether to amplify inputs or weaken them, thereby regulating the flow of sensory and cognitive information. TRN neurons can suppress sensory distractions to support attention (Wimmer *et al.*, 2015). Connectivity between R&C cells and the TRN could provide a mechanism for translating confidence signals into behavioural persistence or disengagement. During the ramp phase, R&C cell modulation of TRN activity may sustain focused thalamocortical gating, which could help to maintain attention and commitment to the task. When R&C cells collapse, the reduced input to

the TRN could release this gating, signalling one to reorient attention and adjust behaviour. In this way, the OFC → TRN pathway may act as a bridge between internal confidence evaluation and the decision of where to focus attention: wait for reward or abandon.

This interpretation aligns well with the activity of ramping unit 298 which was exclusively TRN-optotagged and showed a sharp rise in firing rate upon reward delivery, potentially signalling to the TRN to upregulate attention towards consummatory behaviour (drinking the water reward). On the contrary, during abandonment trials, the collapse in firing activity could signal when to shift attention to the next trial.

4.5 The Pontine Nucleus

The pons receives the third-largest output from the cerebral cortex with approximately 20 million axonal projections from each hemisphere, despite its widespread connectivity to the cortex, the topography, organisation, and function of this area is only marginally defined (Karbaforoushan, Tian & Baker, 2022). Historically, the PN has been viewed as a passive relay center for motor information between the cortex and the cerebellum (Nagao, 2004). Due to this assumption, research has focused on the cerebellum, where computation happens. The cerebellum was typically considered to be an exclusively motor structure; only recently has research established that the area is involved in cognitive processing such as attention, memory, executive function, learning, language, and visuospatial information (Zhang *et al.,* 2023). Thus, the cortico-pontine-cerebellar pathway is not purely for motoric relay but also has a role in higher order cognition.

I found 47 PN-optotagged cells with two of these being R&C cells that project to the PN (figs.28 & 30). Based on the literature, there are three presumptions I would like to make about this OFC → PN network node: first, OFC confidence encoding cells may send movement information downstream via the cortico-pontine-cerebellar pathway, information such as when to abandon the task and move to the left or right port, as well as the necessary motor coordination to execute this. Second, OFC confidence encoding cells could be relaying more complex cognitive signals via the PN to be computed in the cerebellum. What these signals could be, I do not know, this is mere speculation. Third, the PN is not just a relay structure but could have its own specialised roles in cognition. The reason I make this last presumption is because the PN is severely understudied, and to my knowledge no research directly evaluates its role in cognition. As the saying goes, “absence of evidence, is not evidence of absence*”- Carl Sagan* (Feres & Feres, 2023, p1). Therefore, on that premise, I can not rule out the possibility that the OFC → PN projections serve *only* as a motor relay pathway to the cerebellum. In consideration of how densely connected the PN is to the cortex, I am suspicious of its simplicity. More research is clearly needed to evaluate such ponderings, but for the sake of discussion I stand by them.

4.6 Limitations: to do science is to fail time and time again

Until one day you don't! There are numerous reasons why experiments may fail and the most important thing to *understand* is *why* they failed. During my thesis a key setback was that the original implants were unreliable. I had one case where the implant detached and four cases where the neuropixels were corroded by moisture just a few weeks post surgery. After months of trial and error experimenting with headplate designs, Paul Anderson developed new headplates allowing for the recordings to last up to twelve weeks. To prevent moisture-related corrosion, the headplates were custom-designed and 3D-printed with ventilation holes. Silicone protective spray was used to limit moisture and corrosion. Additionally, water bottle access was adjusted to be ground level in order to prevent water leakage into the rats' implants when drinking.

A further limitation was that all animals had failed optotagging experiments except for PMA104. Evidence from histology confirmed that these failures were not due to poor targeting for all animals showed accurate implant placement, pointing instead to viral choice as the cause. Switching to the virus rAAV2/5-hSyn-ChR2(H134R)-eYFP produced good optotagging results in the animal PMA104. Whilst these results for PMA104 were good there were still several limitations. Improper alignment between the ferrule connector and the adapter during mounting can lead to light leakage at the interface; this was visually apparent during the optotagging sessions. Because the animals in this study were freely moving it is likely that this connection was sometimes compromised, decreasing light delivery. One of the optotagging criteria is fidelity; where two thirds of the light pulses had to evoke spikes, if the fiber connection was compromised then this would result in limited light delivery. This inconsistency could lead to only stronger trials eliciting spikes and failure to activate spikes on weaker trials, which could have led to failed fidelity, and thereby false negative results.

Given that a total of 1638 potentially optotagged units did not pass fidelity or modP it is plausible that many of these negative results were due to insufficient light delivery, especially during lower power and frequency trials. False negatives can also occur due to failed virus infection in the target cells, resulting in low or absent CHR2 expression and thus failure to elicit action potentials via light stimulation (Lima *et al.,* 2009). Further, the electrode could have been slightly off target, missing a great portion of OFC cells that did express high CHR2 levels, giving rise to a lower positive rate. On the contrary, it is possible that false positive optotagged units could have occurred due to polysynaptic activation and/or high opsin expression. It is estimated that under good experimental conditions approximately 5-10% of optotagged cells will be false positives (Cohen *et al.*, 2012; Kvistiani *et al.*, 2013). Nonetheless, I suspect this study had a high rate of false negatives and a low rate of false positives due to the stringent criteria that had to be satisfied; this is evident in my results which show ideal distributions for latency, jitter, fidelity, and AUC across conditions (figs.14 &15). The optotagged R&C cells had latencies <0.01s, thus with confidence, I can say they are highly unlikely to be polysynaptic responses (false positives). In fact it is more likely that I am underestimating the number of optotagged cells.

A current limitation of the 2AFC behavioural design is the lack of a control task to distinguish decision-related activity from pure sensory responses. This issue has already been addressed and a new task paradigm developed by Paul Anderson has been used on four animals. This task uses the same three-port setup but without auditory cues; rodents will instead learn to wait for a randomly delivered reward on either the left or right port. This will help isolate neural activity related to reward expectation from stimulus-driven responses and will serve as a reliable control paradigm. Recently, Langdon and Engel (2025) developed a method to separate sensory-driven and decision-related signals by modeling low-dimensional circuit dynamics across neural populations. Adopting this approach could help clarify what activity observed in the behavioural recordings reflected internal decision processes or external sensory input.

The work presented in this thesis is based on the rat model; what we do not know is whether R&C cells exist in humans and encode information in similar ways. This will always be a critique when extrapolating data from animals to humans. Upon the acquisition of reliable confidence results from animal models, human studies regarding R&C computation may be able to take place eventually. One recurring topic in neuroscience is the question of sex differences. In this study, only male rats were used to infer the neural correlates for confidence; this is an obvious limitation. Sex influences neural circuitry, behaviour, and cognitive processing, and inclusive studies on both sexes often lead to more meaningful biological variation (McCarthy *et al.*, 2012; Shansky & Murphy, 2021). The failure to include both sexes in research is a systemic problem that compromises scientific rigor and reinforces sexism. Women deserve equal access to scientific information about themselves; thus, all findings should also be replicated with female models, especially before moving on to human studies.

In sum, the limitations in this thesis were the original implant designs, viral failures, compromised light delivery, lack of a control task, and findings that are limited to one sex. Many of these were eventually mitigated; however, there is room for improvement.

4.7 Future Directions

If you are looking for answers in this thesis, I am afraid you will be sorely disappointed. However, if you are looking to be filled with curiosity and inspiration, then you are in the right place. This thesis presented a myriad of questions and very few definitive answers. Due to time constraints I was unable to live out my wildest scientific dreams regarding this project. As an ode to this loss, in this section I would like to lay out the many unanswered questions I have and propose future directions that could possibly solve them.

Future work should aim to acquire more data from animals during the 2AFC task with successful optotagging in the remaining regions outlined in Mirjam Lambourne's work: MD, PAG, EC, SM, and the OCM. By mapping R&C projections to these regions it will help to understand how confidence is encoded across a network. Once a sufficient amount of data is acquired, the next goal one should make is to understand how R&C neural activity evolves during key epochs; sampling, waiting onset, pre-abandonment and post-abandonment. Then

decipher if these epochs have similar or dissimilar neural states. One could quantify this by using various statistical techniques such as; Pearsons correlation r (Cohen & Kohn, 2011), cosine similarity d (Kriegeskorte, 2008), covariance matrices (Averbeck, Latham, & Pouget, 2006), the Frobenius norm (Golub & Van Loan, 2013), and a participation ratio (Gao *et al.*, 2017). A control dataset will need to be generated to compare these to, a permutation test on random timepoints would be adequate. With this statistical approach, one should be able to examine whether distinct states arise during these epochs, and if they are statistically sound.

To visually assess how these states relate to behaviour one could apply clustering methods; k-means, UMAP or t-SNE, to present the state variance (Macqueen, 1967; Maaten & Hinton, 2008; McInnes *et al.*, 2018). To discover hidden, latent variables one could use dimensionality reduction techniques such as CEBRA, LFADS or MARBLE (Gosztolai *et al.*, 2025; Pandarinath *et al.,* 2018; Schneider *et al.*, 2023). These analyses could help unravel how neural states converge towards decisions such as when to abandon the task or keep waiting. In tandem with these techniques one could use predictive models like elastic net or lasso-regression to determine if R&C cells are the drivers of confidence-dependent waiting in trials (Tibshirani, 1996; Zou & Hastie, 2005). The key question to ask using these predictive models is; can we predict the time of collapse, i.e. how long an animal will wait based on the firing activity during sampling and early wait times? If predictive models can forecast the timing of abandonment this would imply that the "decision" to quit is already embedded in the evolving neural state. Meaning that R&C cells (within a network) encode an internal measure of confidence that directly determines how long an animal is willing to wait. Such a finding would be good evidence that R&C cells play a causal role in confidence-dependent waiting times. This would still need to be confirmed with neuromodulation techniques such as optogenetic inhibition/excitation to determine if the role is sufficient or necessary for confident behaviour. Together, these techniques could help to define how R&C neural patterns are encoding confidence.

Once R&C cell activity is better understood, we can move on to bigger questions: how do R&C cells fit into the wider structure that is the neocortex? In 1979 Vermont Mountcastle flirted with the idea that all regions of the neocortex have the same common architecture. He then elaborated that what distinguishes one region from another is the *inputs* that the region receives, rather than the organisation of circuitry and its subsequent function. He proposed that a single cortical column is the basic computational unit for function (Mountcastle, 1979). Hawkins *et al.* (2019) ran with this idea; the authors were stumped by how a single cortical column could give rise to all cognitive abilities, especially when this was incongruent with already existing hierarchical models of the neocortex. Hawkins' inspiration by Vermont's propositions led to the thousand brains theory of intelligence; the idea that each cortical column of the neocortex builds predictive models of the world based on sensory and motor inputs, with perception and cognition emerging from consensus among these thousands of parallel nodes. Hawkins' theoretical framework is influenced by the discovery of specialised cells that encode specific features of the environment, which together unify to represent a complete percept, cells

such as grid cells, place cells, head direction cells, border cells, and object vector cells (*See references for details*: Deshmukh & Knierim, 2013; Giocomo *et al.*, 2007; Hafting *et al.*, 2005; Lever *et al*., 2009; O'Keefe & Nadel, 1978; O'Keefe & Dostrovsky, 1971; Taube, Muller & Ranck, 1990; Winter, Clark & Taube, 2015). Hawkins' logic is that the neocortex must have feature-specific cells that model the world in a predictive manner, similar to the principles of spatially tuned cells in the hippocampal formation. As such, Hawkins is seeking predictive grid-like cells in the neocortex that help integrate world models. Hawkins' work is primarily focused on object and spatial models in the cortex, however, the theory as a whole relies on predictive networks. If each cortical column builds many predictive models of the world, then there should be neurons that encode the strength and failure of these predictions. I believe R&C cells are compelling candidates for such predictive neurons.

In generalised prediction coding models, the brain maintains an optimal Bayesian estimate of the world by weighing the incoming sensory data against the strength of its predictions, which enables the prediction of any given percept (Kanai *et al.*, 2015; Shipp, 2016). The ramping activity during the waiting period in the 2AFC task could reflect an updating internal prediction that reward is imminent, based on sensory evidence and elapsed time. The collapse phase then signals that the expected reward did not materialise and acts as a prediction error signal that triggers model updating and a shift in behaviour (abandonment of the task). In this context, the R&C cells function as dynamic predictors, with the firing profiles mirroring the strength of the animals' evolving expectations/predictions at each moment. Ramping activity is a neural signature for prediction across the brain, with ramping increasing prior to an upcoming event or reward (Janssen & Shadlen, 200; Khamassi *et al.*, 2008; Schultz, Dayan, & Montague, 1997; Mita *et al.*, 2009; Stalnaker *et al.*, 2014). I presume that R&C cells likely fall within a broader family of prediction-encoding neurons.

If R&C cells are predictive in nature, then how do they encode the strength of confidence, a subjective feeling of certainty? I presume that the very feeling of confidence is an emergent property that occurs downstream of R&C cells. I make this presumption based on Hawkins theory that all cognitive functions emerge from a consensus across predictive models in the neocortex. However, this kind of interpretation opens up a can of worms, that is, the free will discussion… Something worth pondering on, but I will not get into because otherwise this thesis would never end! The reframing of R&C cells as predictive rather than a metacognive emotion (feeling of certainty) is merely speculative, but if one were curious about answering such a question, I would advise them to consider R&C cells through this lens in ongoing future studies. Simply, just keep it in mind.

To substantiate the speculative claims about prediction that are presented here, a complete circuit is required, coupled with a sound understanding of the latent R&C dynamics. Acquired knowledge of what afferent signals R&C cells receive and what behavioural efferent outputs they give would be telling about their function. We can already evince certain roles based on the confirmed projections to the PN, SC, and TRN discovered in this thesis (*See Sections* 4.3, 4.4 & 4.5), however, more data is needed. As I speculated earlier R&C cells may be rich club

neurons, Unit 91 showed strong evidence for this due to its connections with all three target regions, and unit 102 with projections to two targets (Fig.31). Predictive cells are involved in distributed networks, therefore, it would make sense that they are likely rich club neurons, if so, this would be fitting with the data presented in this thesis (Huang *et al.*, 2024). Besides reward prediction, what other contexts would R&C cells need to be examined in to ensure they are involved in predictive-confidence estimates? One could examine how R&C cells respond during tasks involving sensory prediction, timing predictions, and social expectations etcetera. This would help to determine if the R&C phenomena occurs across different modalities, and whether their R&C profiles reflect more abstract confidence predictions. For example, one could investigate why unit 298 had reward-modulated firing (Fig.27). If R&C cells are truly predictive, do they also encode confirmation signals marked by high firing rates, similar to what was observed in 298? What do R&C cells do post outcome? It would be interesting to see how R&C cells respond in different contexts, then investigate if R&C cells are necessary or sufficient for confident predictions via inhibition and/or excitation.

Understanding the neocortex structure to function computations, and how R&C cells play a role in this could help us understand when confidence estimates go awry. Perhaps confidence may not only be a subjective feeling of certainty, but one that is determined by how well sensory inputs align with predictive neurons. Underconfidence or overconfidence may be consequent mismatches between the quality, strength, or number of the afferents that project to R&C cells. Confidence may be an emergent property defined by predictive circuits as well as a metacognitive emotion; feeling of certainty. The R&C cell in nature may be the neural substrate through which the brain quantifies the match between expectation and reality, and signals how to adapt to the world. If so, this could have greater implications for future scientific research. The thousand brains theory proposes that the neocortex is made of thousands of predictive models that compare their outputs and reach consensus between them, giving rise to intelligent behaviour. Hawkins argues that if the neocortex truly operates this way it will not only deepen our understanding of the brain but may also advance AI by building them on similar principles (Hawkins, 2021, *p.*217).

Oscar Wilde once famously wrote in 1989 that “life imitates art” meaning that human perception is shaped by creative representations; art, film, television, the media and so forth… Rather than the other way around (Wilde, 2017). In modern times, life now imitates AI. Today creative representations are now increasingly generated by AI, so much so that the internet is saturated with it. This would not be such a prevalent issue if the AI models were actually correct. People put too much reliance on the correctness of AI, even when it conflicts with their own internal assessments, and this not only distorts our judgements but can lead to poor outcomes (Klingbeil, Grützner, & Schreck, 2024). Large language models are trained to bluff; they are rewarded by giving confident-sounding answers as opposed to honest answers when uncertain, this causes what is known as AI hallucinations: misinformation (Bentegeac *et al.*, 2025; Krishna *et al.*, 2024; Zhao, 2025). Yet, even when trained on perfect data, AI models are still conflicted and uncertain.

Whilst life may imitate AI, the contrary is also true: AI imitates life. Neuroscience and AI have a long interconnected history. A classic example of this is neural networks, machine learning algorithms that are modeled off of the computation of real biological neurons (Macukow, 2016). It is proposed by numerous theoretical scientists that advancements in neuroscience will advance the power of AI and vice versa (Macukow, 2016; Malblestone, Wayne & Kording, 2016; Richards *et al.*, 2019). If we could understand how perception, confidence, and reality align in optimal conditions in humans, we may be able to improve the frameworks for AI confidence assessment and decision-making. Perhaps such AI models could outperform human judgements due to their superior data processing capacity. Building accurate AI models may improve the quality of the information we consume, and this could help inform better decision making on our part. Nonetheless, humans still need to critically evaluate the world independently. It is becoming increasingly difficult to differentiate what content online is true or false, real or fake; as such human confidence assessment needs a revamp. Understanding ourselves better in regards to confidence assessment, may advance the development of targeted therapies, especially for people with disorders where confidence estimation and decision-making are faulty. Solutions could come through in the form of cognitive training, pharmacological intervention, neuromodulation, or other means. It is imperative that we learn to evaluate our own internal assessments of reality more accurately. We are inseparable from technology and are coevolving with it, thus, advancement in one domain will likely advance the other.

4.8 Conclusion

This thesis confirms for the first time in a rat model that cells in the orbitofrontal cortex, including novel ramp and collapse cells, project to the pontine nucleus, superior colliculus, and the thalamic reticular nucleus. This defines a starting point: the first nodes of a confidence network in which to build upon. With further research on ramp and collapse cell activity, especially within a network, we may be able to demystify the black box that is confidence computation in the brain. Enlightenment in this respect may help humanity survive the age of misinformation, defy our own biased heuristics, and assess our own internal reflections of reality more accurately. Or at least, understand ourselves better.

References

Armstrong, J. S., & Plous, S. (1994). The psychology of judgment and decision making. *Journal of Marketing*, *58*(3), 119. https://doi.org/10.2307/1252316

Asadzandi, S., Mojtahedzadeh, R., & Mohammadi, A. (2022). What are the Factors that Enhance Metacognitive Skills in Nursing Students? A Systematic Review. *Iranian Journal of Nursing and Midwifery Research*, *27*(6), 475–484. https://doi.org/10.4103/ijnmr.ijnmr_247_21

Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, *7*(5), 358–366. https://doi.org/10.1038/nrn1888

Banghart, M., Borges, K., Isacoff, E., Trauner, D., & Kramer, R. H. (2004). Light-activated ion channels for remote control of neuronal firing. *Nature Neuroscience*, *7*(12), 1381–1386. https://doi.org/10.1038/nn1356

Basso, M. A., & May, P. J. (2017). Circuits for action and cognition: A view from the superior colliculus. *Annual Review of Vision Science*, *3*(1), 197–226. https://doi.org/10.1146/annurev-vision-102016-061234

Beniaguev, D., Segev, I., & London, M. (2021). Single cortical neurons as deep artificial neural networks. Neuron, 109(17), 2727-2739.e3. https://doi.org/10.1016/j.neuron.2021.07.002

Bentegeac, R., Le Guellec, B., Kuchcinski, G., Amouyel, P., & Hamroun, A. (2025). Token probabilities to mitigate Large Language Models overconfidence in answering medical questions: Quantitative study. *Journal of Medical Internet Research*, *27*, e64348–e64348. https://doi.org/10.2196/64348

Berner, E. S., & Graber, M. L. (2008). Overconfidence as a cause of diagnostic error in medicine. *The American Journal of Medicine*, *121*(5), S2–S23. https://doi.org/10.1016/j.amjmed.2008.01.001

Botadra, R. B., Fang, H., Chen, C. J., Scholl, L., Huang, Y. L., Le, T., ... & Hauck, S. A. (2022) Co-Design of Hardware and Software Systems for Future Integrated Brain-Computer Interfaces.

Boucher, P. O., Wang, T., Carceroni, L., Kane, G., Shenoy, K. V., & Chandrasekaran, C. (2023). Initial conditions combine with sensory evidence to induce decision-related dynamics in premotor cortex. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-41752-2

Brody, C. D. (2003). Timing and neural encoding of somatosensory parametric working memory in macaque prefrontal cortex. *Cerebral Cortex*, *13*(11), 1196–1207. https://doi.org/10.1093/cercor/bhg100

Bussell, J. J., Badman, R. P., Márton, C. D., Bromberg-Martin, E. S., Abbott, L. F., Rajan, K., & Axel, R. (2023). *Representations of the intrinsic value of information in mouse orbitofrontal cortex*. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2023.10.13.562291

Campbell, P. W., Govindaiah, G., & Guido, W. (2024). Development of reciprocal connections between the dorsal lateral geniculate nucleus and the thalamic reticular nucleus. *Neural Development*, *19*(1). https://doi.org/10.1186/s13064-024-00183-5

Carroll, B. J., Sampathkumar, V., Kasthuri, N., & Sherman, S. M. (2022). Layer 5 of cortex innervates the thalamic reticular nucleus in mice. *Proceedings of the National Academy of Sciences*, *119*(38). https://doi.org/10.1073/pnas.2205209119

Cavanaugh, J., McAlonan, K., & Wurtz, R. H. (2020). Organization of corollary discharge neurons in monkey medial dorsal thalamus. *The Journal of Neuroscience*, *40*(33), 6367–6378. https://doi.org/10.1523/jneurosci.2344-19.2020

Churchland, M. M., Santhanam, G., & Shenoy, K. V. (2006). Preparatory activity in premotor and motor cortex reflects the speed of the upcoming reach. *Journal of Neurophysiology*, *96*(6), 3130–3146. https://doi.org/10.1152/jn.00307.2006

Cohen, J. Y., Haesler, S., Vong, L., Lowell, B. B., & Uchida, N. (2012). Neuron-type-specific signals for reward and punishment in the ventral tegmental area. *Nature*, *482*(7383), 85–88. https://doi.org/10.1038/nature10754

Cohen, M. R., & Kohn, A. (2011). Measuring and interpreting neuronal correlations. *Nature Neuroscience*, *14*(7), 811–819. https://doi.org/10.1038/nn.2842

Cui, P., Song, K., Mariatos-Metaxas, D., Isla, A. G., Femenia, T., Lazaridis, I., Meletis, K., Kumar, A., & Kardamakis, A. A. (2025). Recurrent circuits encode de novo visual center-surround computations in the mouse superior colliculus. *PLOS Biology*, *23*(10), e3003414. https://doi.org/10.1371/journal.pbio.3003414

Curtis, D., & Andersen, P. (2000a). John carew eccles 1903-1997. *Historical Records of Australian Science*, *13*(4), 439. https://doi.org/10.1071/hr0011340439

Deisseroth, K. (2010). Optogenetics. *Nature Methods*, *8*(1), 26–29. https://doi.org/10.1038/nmeth.f.324

Deisseroth, K., Feng, G., Majewska, A. K., Miesenböck, G., Ting, A., & Schnitzer, M. J. (2006). Next-Generation optical technologies for illuminating genetically targeted brain circuits. *The Journal of Neuroscience*, *26*(41), 10380–10386. https://doi.org/10.1523/jneurosci.3863-06.2006

De Neys, W., Cromheeke, S., & Osman, M. (2011). Biased but in doubt: Conflict and decision confidence. *PLoS ONE*, *6*(1), e15954. https://doi.org/10.1371/journal.pone.0015954

Deshmukh, S. S., & Knierim, J. J. (2013). Influence of local objects on hippocampal representations: Landmark vectors and memory. *Hippocampus*, *23*(4), 253–267. https://doi.org/10.1002/hipo.22101

Dorris, M. C., Paré, M., & Munoz, D. P. (1997). Neuronal activity in monkey superior colliculus related to the initiation of saccadic eye movements. *The Journal of Neuroscience*, *17*(21), 8566–8579. https://doi.org/10.1523/jneurosci.17-21-08566.1997

Duffy, B. (2018). *The perils of perception: Why we're wrong about nearly everything*. Atlantic Books.

Endo, D., Kobayashi, R., Bartolo, R., Averbeck, B. B., Sugase-Miyamoto, Y., Hayashi, K., Kawano, K., Richmond, B. J., & Shinomoto, S. (2021). A convolutional neural network for estimating synaptic connectivity from spike trains. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-91244-w

English, D. F., McKenzie, S., Evans, T., Kim, K., Yoon, E., & Buzsski, G. (2018). Pyramidal cell-interneuron circuit architecture and dynamics in hippocampal networks. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3155904

Ernst, M., & Paulus, M. P. (2005). Neurobiology of decision making: A selective review from a neurocognitive and clinical perspective. *Biological Psychiatry*, *58*(8), 597–604. https://doi.org/10.1016/j.biopsych.2005.06.004

Fatahi, Z., Haghparast, A., Khani, A., & Kermani, M. (2018). Functional connectivity between anterior cingulate cortex and orbitofrontal cortex during value-based decision making.

Neurobiology of Learning and Memory, *147*, 74–78. https://doi.org/10.1016/j.nlm.2017.11.014

Fellows, L. K. (2011). Orbitofrontal contributions to value‐based decision making: Evidence from humans with frontal lobe damage. *Annals of the New York Academy of Sciences*, *1239*(1), 51–58. https://doi.org/10.1111/j.1749-6632.2011.06229.x

Feres, M., & Feres, M. F. N. (2023). Absence of evidence is not evidence of absence. *Journal of Applied Oral Science*, *31*. https://doi.org/10.1590/1678-7757-2023-ed001

Fernandez-Vargas, J., Tremmel, C., Valeriani, D., Bhattacharyya, S., Cinel, C., Citi, L., & Poli, R. (2021). Subject- and task-independent neural correlates and prediction of decision confidence in perceptual decision making. *Journal of Neural Engineering*, *18*(4), 046055. https://doi.org/10.1088/1741-2552/abf2e4

Fleming, S. M., & Dolan, R. J. (2010). Effects of loss aversion on post-decision wagering: Implications for measures of awareness. *Consciousness and Cognition*, *19*(1), 352–363. https://doi.org/10.1016/j.concog.2009.11.002

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00443

Filevich, E., Koß, C., & Faivre, N. (2020). Response-Related signals increase confidence but not metacognitive performance. *Eneuro*, *7*(3), ENEURO.0326-19.2020. https://doi.org/10.1523/eneuro.0326-19.2020

Foote, A. L., & Crystal, J. D. (2007). Metacognition in the rat. *Current Biology*, *17*(6), 551–555. https://doi.org/10.1016/j.cub.2007.01.061

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2213–2223. https://doi.org/10.1098/rstb.2012.0123

Fuster, J. M. (2015). Anatomy of the prefrontal cortex. In *The Prefrontal Cortex* (pp. 9–62). Elsevier. https://doi.org/10.1016/b978-0-12-407815-4.00002-7

Gao, P., Trautmann, E., Yu, B., Santhanam, G., Ryu, S., Shenoy, K., & Ganguli, S. (2017). *A theory of multineuronal dimensionality, dynamics and measurement*. Cold Spring Harbor Laboratory. https://doi.org/10.1101/214262

Giocomo, L. M., Zilli, E. A., Fransén, E., & Hasselmo, M. E. (2007). Temporal frequency of subthreshold oscillations scales with entorhinal grid cell field spacing. *Science*, *315*(5819), 1719–1722. https://doi.org/10.1126/science.1139207

Goodall, J. (1964). Tool-Using and aimed throwing in a community of free-living chimpanzees. *Nature*, *201*(4926), 1264–1266. https://doi.org/10.1038/2011264a0

Goodie, A. S. (2003). The effects of control on betting: Paradoxical betting on items of high confidence with low value. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(4), 598–610. https://doi.org/10.1037/0278-7393.29.4.598

Golub, G. H., & Van Loan, C. F. (2013). *Matrix computations 4th edition the johns hopkins university press.* Baltimore, MD.

Gosztolai, A., Peach, R. L., Arnaudon, A., Barahona, M., & Vandergheynst, P. (2025). MARBLE: Interpretable representations of neural population dynamics using geometric deep learning. *Nature Methods*, *22*(3), 612–620. https://doi.org/10.1038/s41592-024-02582-2

Grayson, D. S., Ray, S., Carpenter, S., Iyer, S., Dias, T. G. C., Stevens, C., Nigg, J. T., & Fair, D. A. (2014). Structural and functional rich club organization of the brain in children and adults. *PLoS ONE*, *9*(2), e88297. https://doi.org/10.1371/journal.pone.0088297

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806. https://doi.org/10.1038/nature03721

Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences*, *98*(9), 5359–5362. https://doi.org/10.1073/pnas.071600998

Hardie, R. (2002). Faculty Opinions recommendation of Selective photostimulation of genetically chARGed neurons. *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature*. https://doi.org/10.3410/f.1003851.42255

Harlow, J. M. (1999). Passage of an iron rod through the head. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *11*(2), 281–283. https://doi.org/10.1176/jnp.11.2.281

Harun, R., Jun, E., Park, H. H., Ganupuru, P., Goldring, A. B., & Hanks, T. D. (2020). Timescales of evidence evaluation for decision making and associated confidence judgments are adapted to task demands. *Frontiers in Neuroscience*, *14*. https://doi.org/10.3389/fnins.2020.00826

Hawkins, J. (2021). *A thousand brains: a new theory of intelligence*. Hachette UK.

Hawkins, J., Lewis, M., Klukas, M., Purdy, S., & Ahmad, S. (2019). A framework for intelligence and cortical function based on grid cells in the neocortex. *Frontiers in Neural Circuits*, *12*. https://doi.org/10.3389/fncir.2018.00121

Healy, P. J., & Moore, D. A. (2007). The trouble with overconfidence. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1001821

Heereman, J., & Walla, P. (2011). Stress, uncertainty and decision confidence. *Applied Psychophysiology and Biofeedback*, *36*(4), 273–279. https://doi.org/10.1007/s10484-011-9167-9

Hellmann, S., Zehetleitner, M., & Rausch, M. (2024). Confidence is influenced by evidence accumulation time in dynamical decision models. *Computational Brain & Behavior*, *7*(3), 287–313. https://doi.org/10.1007/s42113-024-00205-9

Hirokawa, J., Vaughan, A., Masset, P., Ott, T., & Kepecs, A. (2019). Frontal cortex neuron types categorically encode single decision variables. *Nature*, *576*(7787), 446–451. https://doi.org/10.1038/s41586-019-1816-9

Huang, Y. T., Wu, C.-T., Fang, Y.-X. M., Fu, C.-K., Koike, S., & Chao, Z. C. (2024). Crossmodal hierarchical predictive coding for audiovisual sequences in the human brain. *Communications Biology*, *7*(1). https://doi.org/10.1038/s42003-024-06677-6

Janssen, P., & Shadlen, M. N. (2005). A representation of the hazard rate of elapsed time in macaque area LIP. *Nature Neuroscience*, *8*(2), 234–241. https://doi.org/10.1038/nn1386

Jazayeri, M., & Shadlen, M. N. (2015). A neural mechanism for sensing and reproducing a time interval. *Current Biology*, *25*(20), 2599–2609. https://doi.org/10.1016/j.cub.2015.08.038

Joo, H. R., Liang, H., Chung, J. E., Geaghan-Breiner, C., Fan, J. L., Nachman, B. P., Kepecs, A., & Frank, L. M. (2021). Rats use memory confidence to guide decisions. *Current Biology*, *31*(20), 4571-4583.e4. https://doi.org/10.1016/j.cub.2021.08.013

Jun, E. J., Bautista, A. R., Nunez, M. D., Allen, D. C., Tak, J. H., Alvarez, E., & Basso, M. A. (2021). Causal role for the primate superior colliculus in the computation of evidence for

perceptual decisions. *Nature Neuroscience*, *24*(8), 1121–1131. https://doi.org/10.1038/s41593-021-00878-6

Junge, W. (2002). Faculty Opinions recommendation of Channelrhodopsin-1: A light-gated proton channel in green algae. *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature*. https://doi.org/10.3410/f.1008423.105707

Kahneman, D. (2011). *Thinking, fast and slow*. p32. Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251. https://doi.org/10.1037/h0034747

Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *370*(1668), 20140169. https://doi.org/10.1098/rstb.2014.0169

Karbasforoushan, H., Tian, R., & Baker, J. (2022). There is a topographic organization in human cortico-pontine connectivity. *Brain Communications*, *4*(2). https://doi.org/10.1093/braincomms/fcac047

Kaufman, M. T., Churchland, M. M., Ryu, S. I., & Shenoy, K. V. (2014). Cortical activity in the null space: Permitting preparation without movement. *Nature Neuroscience*, *17*(3), 440–448. https://doi.org/10.1038/nn.3643

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1322–1337. https://doi.org/10.1098/rstb.2012.0037

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227–231. https://doi.org/10.1038/nature07200

Khamassi, M., Mulder, A. B., Tabuchi, E., Douchamps, V., & Wiener, S. I. (2008). Anticipatory reward signals in ventral striatal neurons of behaving rats. *European Journal of Neuroscience*, *28*(9), 1849–1866. https://doi.org/10.1111/j.1460-9568.2008.06480.x

Kim, Y. J., Packer, O., Macrina, T., Pollreisz, A., Curcio, C. A., Lee, K., Kemnitz, N., Ih, D., Nguyen, T., Lu, R., Popovych, S., Halageri, A., Bae, J. A., Strout, J., Gerhard, S., Smith, R. G., Martin, P. R., Grünert, U., & Dacey, D. M. (2025). *Visual pathway origins: An electron microscopic volume for connectomic analysis of the human foveal retina*. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2025.04.05.647403

Klingbeil, A., Grützner, C., & Schreck, P. (2024). Trust and reliance on AI — An experimental study on the extent and costs of overreliance on AI. *Computers in Human Behavior*, *160*, 108352. https://doi.org/10.1016/j.chb.2024.108352

Kvitsiani, D., Ranade, S., Hangya, B., Taniguchi, H., Huang, J. Z., & Kepecs, A. (2013). Distinct behavioural and network correlates of two interneuron types in prefrontal cortex. *Nature*, *498*(7454), 363–366. https://doi.org/10.1038/nature12176

Ko, H., Hofer, S. B., Pichler, B., Buchanan, K. A., Sjöström, P. J., & Mrsic-Flogel, T. D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature*, *473*(7345), 87–91. https://doi.org/10.1038/nature09880

Ko, Y. H., Zhou, A., Niessen, E., Stahl, J., Weiss, P. H., Hester, R., Bode, S., & Feuerriegel, D. (2024). Neural correlates of confidence during decision formation in a perceptual judgment task. *Cortex*, *173*, 248–262. https://doi.org/10.1016/j.cortex.2024.01.006

Kobak, D., Brendel, W., Constantinidis, C., Feierstein, C. E., Kepecs, A., Mainen, Z. F., Qi, X.-L., Romo, R., Uchida, N., & Machens, C. K. (2016). Demixed principal component analysis of neural population data. *eLife*, *5*. https://doi.org/10.7554/elife.10989

Kobayashi, R., Kurita, S., Kurth, A., Kitano, K., Mizuseki, K., Diesmann, M., Richmond, B. J., & Shinomoto, S. (2019). Reconstructing neuronal circuitry from parallel spike trains. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-12225-2

Kobayashi, R., & Shinomoto, S. (2024). Inference of monosynaptic connections from parallel spike trains: A review. *Neuroscience Research*. https://doi.org/10.1016/j.neures.2024.07.006

Krishna, S., Bhambra, N., Bleakney, R., & Bhayana, R. (2024). Evaluation of reliability, repeatability, robustness, and confidence of GPT-3.5 And GPT-4 on a radiology board–style examination. *Radiology*, *311*(2). https://doi.org/10.1148/radiol.232715

Lagler, M., Anderson, P., Nathanson, B.-O., Masset, P., Ott, T., Christensen, A., Malagon-Vina, H., Hauer, R., Solano Mateos, M., Kepecs, A., & Klausberger, T. (n.d.). *Orbitofrontal ramp-and-collapse cells for post-decision confidence and metacognition* [Unpublished manuscript].

Lak, A., Costa, G. M., Romberg, E., Koulakov, A. A., Mainen, Z. F., & Kepecs, A. (2014). Orbitofrontal cortex is required for optimal waiting based on decision confidence. *Neuron*, *84*(1), 190–201. https://doi.org/10.1016/j.neuron.2014.08.039

Lak, A., Hueske, E., Hirokawa, J., Masset, P., Ott, T., Urai, A. E., Donner, T. H., Carandini, M., Tonegawa, S., Uchida, N., & Kepecs, A. (2020). Reinforcement biases subsequent perceptual decisions when confidence is low, a widespread behavioral phenomenon. *eLife*, *9*. https://doi.org/10.7554/elife.49834

Lak, A., Okun, M., Moss, M. M., Gurnani, H., Farrell, K., Wells, M. J., Reddy, C. B., Kepecs, A., Harris, K. D., & Carandini, M. (2020). Dopaminergic and prefrontal basis of learning

from sensory confidence and reward value. *Neuron*, *105*(4), 700-711.e6. https://doi.org/10.1016/j.neuron.2019.11.018

Langdon, C., & Engel, T. A. (2025). Latent circuit inference from heterogeneous neural responses during cognitive tasks. *Nature Neuroscience*, *28*(3), 665–675. https://doi.org/10.1038/s41593-025-01869-7

Lerman, K., Yan, X., & Wu, X.-Z. (2016). The "majority illusion" in social networks. *PLOS ONE*, *11*(2), e0147617. https://doi.org/10.1371/journal.pone.0147617

Lever, C., Burton, S., Jeewajee, A., O'Keefe, J., & Burgess, N. (2009). Boundary vector cells in the subiculum of the hippocampal formation. *The Journal of Neuroscience*, *29*(31), 9771–9777. https://doi.org/10.1523/jneurosci.1319-09.2009

Li, C., Zhang, Y., Wang, T., Xu, X., Wang, Q., Xu, B., & Cui, H. (2020). Generative decoding of intracortical neuronal signals for online control of robotic arm to intercept moving objects. *Journal of Physics: Conference Series*, *1576*(1), 012057. https://doi.org/10.1088/1742-6596/1576/1/012057

Li, L., & Wang, D. (2021). Adaptive neurons compute confidence in a decision network. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-01523-9

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). *Brain*, *106*(3), 623–642. https://doi.org/10.1093/brain/106.3.623

Lima, S. Q., Hromádka, T., Znamenskiy, P., & Zador, A. M. (2009). PINP: A new method of tagging neuronal populations for identification during in vivo electrophysiological recording. *PLoS ONE*, *4*(7), e6099. https://doi.org/10.1371/journal.pone.0006099

Liu, Z. (2024). The asymmetric impact of decision-making confidence on regret and relief. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1365743

Maaten, L.V., & Hinton, G.E. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9, 2579-2605.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations.

Macukow, B. (2016). Neural networks – state of art, brief history, basic models and architecture. In *Lecture Notes in Computer Science* (pp. 3–14). Springer International Publishing. https://doi.org/10.1007/978-3-319-45378-1_1

Marblestone, A. H., Wayne, G., & Kording, K. P. (2016). Toward an integration of deep learning and neuroscience. *Frontiers in Computational Neuroscience*, *10*. https://doi.org/10.3389/fncom.2016.00094

Masset, P., Ott, T., Lak, A., Hirokawa, J., & Kepecs, A. (2020). behaviour- and modality-general representation of confidence in orbitofrontal cortex. *Cell*, *182*(1), 112-126.e18. https://doi.org/10.1016/j.cell.2020.05.022

Massot, C., Jagadisan, U. K., & Gandhi, N. J. (2019). Sensorimotor transformation elicits systematic patterns of activity along the dorsoventral extent of the superior colliculus in the macaque monkey. *Communications Biology*, *2*(1). https://doi.org/10.1038/s42003-019-0527-y

Mathis, A., Mamidanna, P., Cury, K. M., Abe, T., Murthy, V. N., Mathis, M. W., & Bethge, M. (2018). DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, *21*(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

Maunsell, J. H. R. (2015). Neuronal mechanisms of visual attention. *Annual Review of Vision Science*, *1*(1), 373–391. https://doi.org/10.1146/annurev-vision-082114-035431

McAlonan, K., Cavanaugh, J., & Wurtz, R. H. (2006). Attentional modulation of thalamic reticular neurons. *The Journal of Neuroscience*, *26*(16), 4444–4450. https://doi.org/10.1523/jneurosci.5602-05.2006

McCarthy, M. M., Arnold, A. P., Ball, G. F., Blaustein, J. D., & De Vries, Geert. J. (2012). Sex differences in the brain: the not so inconvenient truth. *The Journal of Neuroscience*, *32*(7), 2241–2247. https://doi.org/10.1523/jneurosci.5372-11.2012

McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018a). UMAP: uniform manifold approximation and projection. *Journal of Open Source Software*, *3*(29), 861. https://doi.org/10.21105/joss.00861

Mecklenbrauck, F., Gruber, M., Siestrup, S., Zahedi, A., Grotegerd, D., Mauritz, M., Trempler, I., Dannlowski, U., & Schubotz, R. I. (2023). The significance of structural rich club hubs for the processing of hierarchical stimuli. *Human Brain Mapping*, *45*(4). https://doi.org/10.1002/hbm.26543

Merchant, H., & Averbeck, B. B. (2017). The computational and neural basis of rhythmic timing in medial premotor cortex. *The Journal of Neuroscience*, *37*(17), 4552–4564. https://doi.org/10.1523/jneurosci.0367-17.2017

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as bayesian probability: From neural origins to behaviour. *Neuron*, *88*(1), 78–92. https://doi.org/10.1016/j.neuron.2015.09.039

Mita, A., Mushiake, H., Shima, K., Matsuzaka, Y., & Tanji, J. (2009). Interval time coding by neurons in the presupplementary and supplementary motor areas. *Nature Neuroscience*, *12*(4), 502–507. https://doi.org/10.1038/nn.2272

Morelli, M., Casagrande, M., & Forte, G. (2021). Decision making: A theoretical review. *Integrative Psychological and behavioural Science*, *56*(3), 609–629. https://doi.org/10.1007/s12124-021-09669-x

Moreno, L., Briñol, P., & Petty, R. E. (2021). Metacognitive confidence can increase but also decrease performance in academic settings. *Metacognition and Learning*, *17*(1), 139–165. https://doi.org/10.1007/s11409-021-09270-y

Mountcastle, V. B. (1979). An organizing principle for cerebral function. *The mindful brain*, 7-50.

Murphy, D. H., Rhodes, M. G., & Castel, A. D. (2024). Age-related differences in metacognitive reactivity in younger and older adults. *Metacognition and Learning*, *19*(3), 863–877. https://doi.org/10.1007/s11409-024-09391-0

Nagao, S. (2004). Pontine nuclei-mediated cerebello-cerebral interactions and its functional role. *The Cerebellum*, *3*(1), 11–15. https://doi.org/10.1080/14734220310012181

Nagel, G., Ollig, D., Fuhrmann, M., Kateriya, S., Musti, A. M., Bamberg, E., & Hegemann, P. (2002a). Channelrhodopsin-1: A light-gated proton channel in green algae. *Science*, *296*(5577), 2395–2398. https://doi.org/10.1126/science.1072068

Nozaki, D. (2024). Level of decision confidence shapes motor memory. *Trends in Cognitive Sciences*, *28*(9), 786–788. https://doi.org/10.1016/j.tics.2024.07.010

Ogasa, K., Yokoi, A., Okazawa, G., Nishigaki, M., Hirashima, M., & Hagura, N. (2024). Decision uncertainty as a context for motor memory. *Nature Human Behaviour*, *8*(9), 1738–1751. https://doi.org/10.1038/s41562-024-01911-x

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*(1), 171–175. https://doi.org/10.1016/0006-8993(71)90358-1

O'keefe, J., & Nadel, L. (1978). The hippocampus as a cognitive map. *Oxford university press*.

Pachitariu, M., Sridhar, S., Pennington, J., & Stringer, C. (2024). Spike sorting with Kilosort4. *Nature methods*, *21*(5), 914-921.

Pandarinath, C., O'Shea, D. J., Collins, J., Jozefowicz, R., Stavisky, S. D., Kao, J. C., Trautmann, E. M., Kaufman, M. T., Ryu, S. I., Hochberg, L. R., Henderson, J. M., Shenoy, K. V., Abbott, L. F., & Sussillo, D. (2018). Inferring single-trial neural population dynamics using sequential auto-encoders. *Nature Methods*, *15*(10), 805–815. https://doi.org/10.1038/s41592-018-0109-9

Papp, E. A., Leergaard, T. B., Calabrese, E., Johnson, G. A., & Bjaalie, J. G. (2014). Waxholm Space atlas of the Sprague Dawley rat brain. *NeuroImage*, *97*, 374–386. https://doi.org/10.1016/j.neuroimage.2014.04.001

Passage of an iron rod through the head. (1848). *The Boston Medical and Surgical Journal*, *39*(20), 389–393. https://doi.org/10.1056/nejm184812130392001Paxinos, G., & Watson, C. (2005). The rat brain in stereotaxic coordinates (5th ed.). Elsevier Academic Press.

Perin, R., Berger, T. K., & Markram, H. (2011a). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences*, *108*(13), 5419–5424. https://doi.org/10.1073/pnas.1016051108

Perry, C. J., & Barron, A. B. (2013). Honey bees selectively avoid difficult choices. *Proceedings of the National Academy of Sciences*, *110*(47), 19155–19159. https://doi.org/10.1073/pnas.1314571110

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience*, *10*(2), 257–261. https://doi.org/10.1038/nn1840

Pierre, J. M. (2024). Conspiracies gone wild: A psychiatric perspective on conspiracy theory belief, mental illness, and the potential for lone actor ideological violence. *Terrorism and Political Violence*, *37*(4), 550–569. https://doi.org/10.1080/09546553.2024.2329079

Rahnev, D. (2025). A comprehensive assessment of current methods for measuring metacognition. *Nature Communications*, *16*(1). https://doi.org/10.1038/s41467-025-56117-0

Ran, X., Chen, W., Yvert, B., & Zhang, S. (2022). A hybrid autoencoder framework of dimensionality reduction for brain-computer interface decoding. *Computers in Biology and Medicine*, *148*, 105871. https://doi.org/10.1016/j.compbiomed.2022.105871

Ray, P., Reddy, S. S., & Banerjee, T. (2021). Various dimension reduction techniques for high dimensional data analysis: A review. *Artificial Intelligence Review*, *54*(5), 3473–3515. https://doi.org/10.1007/s10462-020-09928-0

Redgrave, P. (2010). Interactions between the midbrain superior colliculus and the basal ganglia. *Frontiers in Neuroanatomy*, *4*. https://doi.org/10.3389/fnana.2010.00132

Ren, N., Ito, S., Hafizi, H., Beggs, J. M., & Stevenson, I. H. (2020). Model-based detection of putative synaptic connections from spike recordings with latency and type constraints. *Journal of Neurophysiology*, *124*(6), 1588–1604. https://doi.org/10.1152/jn.00066.2020

Rodriguez, N., Bollen, J., & Ahn, Y.-Y. (2016). Collective dynamics of belief evolution under cognitive coherence and social conformity. *PLOS ONE*, *11*(11), e0165910. https://doi.org/10.1371/journal.pone.0165910

Roesch, M. R., Taylor, A. R., & Schoenbaum, G. (2006). Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. *Neuron*, *51*(4), 509–520. https://doi.org/10.1016/j.neuron.2006.06.027

Rolls, E. T., Cheng, W., & Feng, J. (2020). The orbitofrontal cortex: Reward, emotion and depression. *Brain Communications*, *2*(2). https://doi.org/10.1093/braincomms/fcaa196

Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a statistical computation in the human sense of confidence. *Neuron*, *90*(3), 499–506. https://doi.org/10.1016/j.neuron.2016.03.025

Sapolsky, R. (2024). Life without free will: Does it preclude possibilities? *Possibility Studies & Society*, *2*(3), 272–281. https://doi.org/10.1177/27538699241284330

Savitzky, Abraham., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, *36*(8), 1627–1639. https://doi.org/10.1021/ac60214a047

Schneider, S., Lee, J. H., & Mathis, M. W. (2023). Learnable latent embeddings for joint behavioural and neural analysis. *Nature*, *617*(7960), 360–368. https://doi.org/10.1038/s41586-023-06031-6

Schnitzler, A., & Wojtecki, L. (2012). Faculty Opinions recommendation of Lesion mapping of cognitive control and value-based decision making in the prefrontal cortex. *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature*. https://doi.org/10.3410/f.717956183.793461058

Schoenbaum, G., Roesch, M. R., Stalnaker, T. A., & Takahashi, Y. K. (2009). A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nature Reviews Neuroscience*, *10*(12), 885–892. https://doi.org/10.1038/nrn2753

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Shansky, R. M., & Murphy, A. Z. (2021). Considering sex as a biological variable will require a global shift in science culture. *Nature Neuroscience*, *24*(4), 457–464. https://doi.org/10.1038/s41593-021-00806-8

Sheng, K., Zhang, S., Zhang, S., He, Y., Beau, M., Qu, P., Liu, X., Zhang, Y., Ma, L., & Du, K. (2022). *Reconstructing monosynaptic connectivity from in vivo spike trains using deep domain-adaptive matching* [Preprint]. bioRxiv. https://doi.org/10.1101/2022.10.03.510694

Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology, 7.* https://doi.org/10.3389/fpsyg.2016.01792

Sigmoid Function - An overview. (n.d.). ScienceDirect Topics. Retrieved October 21, 2025, from https://www.sciencedirect.com/topics/computer-science/sigmoid-function#

Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences*, *13*(9), 389–396. https://doi.org/10.1016/j.tics.2009.06.009

Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor ,R., & Erb, L. (1995). The uncertain response in the bottlenosed dolphin (Tursiops truncatus). *Journal of Experimental Psychology: General*, *124*(4), 391–408. https://doi.org/10.1037//0096-3445.124.4.391

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., & Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, *3*(3), e68. https://doi.org/10.1371/journal.pbio.0030068

Sonuga‐Barke, E. J. S., Cortese, S., Fairchild, G., & Stringaris, A. (2015). Annual Research Review: Transdiagnostic neuroscience of child and adolescent mental disorders – differentiating decision making in attention‐deficit/hyperactivity disorder, conduct disorder, depression, and anxiety. *Journal of Child Psychology and Psychiatry*, *57*(3), 321–349. https://doi.org/10.1111/jcpp.12496

Sosa, J. L. R., Buonomano, D., & Izquierdo, A. (2021). The orbitofrontal cortex in temporal cognition. *behavioural Neuroscience*, *135*(2), 154–164. https://doi.org/10.1037/bne0000430

Stalnaker, T. A., Cooch, N. K., McDannald, M. A., Liu, T.-L., Wied, H., & Schoenbaum, G. (2014). Orbitofrontal neurons infer the value and identity of predicted outcomes. *Nature Communications*, *5*(1). https://doi.org/10.1038/ncomms4926

Steinmetz, N. A., Aydin, C., Lebedeva, A., Okun, M., Pachitariu, M., Bauza, M., Beau, M., Bhagat, J., Böhm, C., Broux, M., Chen, S., Colonell, J., Gardner, R. J., Karsh, B., Kloosterman, F., Kostadinov, D., Mora-Lopez, C., O'Callaghan, J., Park, J., … Harris, T. D. (2021). Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings. *Science*, *372*(6539). https://doi.org/10.1126/science.abf4588

Stevenson, I. H. (2023). Circumstantial evidence and explanatory models for synapses in large-scale spike recordings. *Neurons, behaviour, Data Analysis, and Theory*. https://doi.org/10.51628/001c.90831

Stevenson, I. H., & Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nature Neuroscience*, *14*(2), 139–142. https://doi.org/10.1038/nn.2731

Stevenson, I. H., Rebesco, J. M., Miller, L. E., & Körding, K. P. (2008). Inferring functional connections between neurons. *Current Opinion in Neurobiology*, *18*(6), 582–588. https://doi.org/10.1016/j.conb.2008.11.005

Stolyarova, A., & Izquierdo, A. (2017). Complementary contributions of basolateral amygdala and orbitofrontal cortex to value learning under uncertainty. eLife, 6. https://doi.org/10.7554/elife.27483

Swadlow, H. A. (1995). Influence of VPM afferents on putative inhibitory interneurons in S1 of the awake rabbit: Evidence from cross-correlation, microstimulation, and latencies to peripheral sensory stimulation. *Journal of Neurophysiology*, *73*(4), 1584–1599. https://doi.org/10.1152/jn.1995.73.4.1584

Takai, Y. (2005). Faculty Opinions recommendation of Millisecond-timescale, genetically targeted optical control of neural activity. *Faculty Opinions – Post-Publication Peer Review of the Biomedical Literature*. https://doi.org/10.3410/f.1026686.346497

Taube, J., Muller, R., & Ranck, J., Jr. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. II. Effects of environmental manipulations. *The Journal of Neuroscience*, *10*(2), 436–447. https://doi.org/10.1523/jneurosci.10-02-00436.1990

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Thomas, A., Yang, W., Wang, C., Tipparaju, S. L., Chen, G., Sullivan, B., Swiekatowski, K., Tatam, M., Gerfen, C., & Li, N. (2023). Superior colliculus bidirectionally modulates choice activity in frontal cortex. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-43252-9

Van den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence. *eLife*, *5*. https://doi.org/10.7554/elife.12192

Verasztó, C., Jasek, S., Gühmann, M., Shahidi, R., Ueda, N., Beard, J. D., Mendes, S., Heinz, K., Bezares-Calderón, L. A., Williams, E., & Jékely, G. (2020). *Whole-animal connectome and cell-type complement of the three-segmentedPlatynereis dumeriliilarva*. Cold Spring Harbor Laboratory. https://doi.org/10.1101/2020.08.21.260984

Wilde, O. (2017). *Selected poems of Oscar Wilde*. Createspace Independent Publishing Platform.

Wimmer, R. D., Schmitt, L. I., Davidson, T. J., Nakajima, M., Deisseroth, K., & Halassa, M. M. (2015a). Thalamic control of sensory selection in divided attention. *Nature*, *526*(7575), 705–709. https://doi.org/10.1038/nature15398

Winter, S. S., Clark, B. J., & Taube, J. S. (2015). Disruption of the head direction cell network impairs the parahippocampal grid cell signal. *Science*, *347*(6224), 870–874. https://doi.org/10.1126/science.1259591

White JG, Southgate E, Thomson JN, Brenner S. The structure of the nervous system of the nematodeCaenorhabditis elegans. (1986a). Philosophical Transactions of the Royal Society of London. B, Biological Sciences, 314(1165), 1–340. https://doi.org/10.1098/rstb.1986.0056

Xue, K., Zheng, Y., Rafiei, F., & Rahnev, D. (2023a). The timing of confidence computations in human prefrontal cortex. *Cortex*, *168*, 167–175. https://doi.org/10.1016/j.cortex.2023.08.009

Xue, K., Zheng, Y., Rafiei, F., & Rahnev, D. (2023b). The timing of confidence computations in human prefrontal cortex. *Cortex*, *168*, 167–175. https://doi.org/10.1016/j.cortex.2023.08.009

Zemelman, B. V., Nesnas, N., Lee, G. A., & Miesenböck, G. (2003). Photochemical gating of heterologous ion channels: Remote control over genetically designated populations of neurons. *Proceedings of the National Academy of Sciences*, *100*(3), 1352–1357. https://doi.org/10.1073/pnas.242738899

Zhang, P., Duan, L., Ou, Y., Ling, Q., Cao, L., Qian, H., Zhang, J., Wang, J., & Yuan, X. (2023). The cerebellum and cognitive neural networks. *Frontiers in Human Neuroscience*, *17*. https://doi.org/10.3389/fnhum.2023.1197459

Zhao, C. (2025, October 28). AI hallucinates because it's trained to fake answers it doesn't know. *American Association for the Advancement of Science*. https://www.science.org/content/article/ai-hallucinates-because-it-s-trained-fake-answers-it-doesn-t-know

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x