Neural Dynamics of Numerical Cognition: Coding, Maintenance, and Interference in the Primate Frontoparietal Network

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Tobias Machts

aus Gotha

Tübingen

2026

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	30.07.2026
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter/-in:	Prof. Dr. Andreas Nieder
2. Berichterstatter/-in:	Prof. Dr. Uwe Ilg
3. Berichterstatter/-in:	Prof. Dr. Gregor Rainer

Erklärung / Declaration:

Ich erkläre, dass ich die zur Promotion eingereichte Arbeit mit dem Titel „Neural Dynamics of Numerical Cognition: Coding, Maintenance, and Interference in the Primate Frontoparietal Network" selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

KI-basierte Sprachassistenztools wurden ausschließlich zur sprachlichen Überarbeitung und Verbesserung der Verständlichkeit eingesetzt. Der Autor übernimmt die volle Verantwortung für alle wissenschaftlichen Inhalte, Analysen und Schlussfolgerungen.

I hereby declare that I have produced the work entitled "Neural Dynamics of Numerical Cognition: Coding, Maintenance, and Interference in the Primate Frontoparietal Network", submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

AI-based language assistance tools were used exclusively for linguistic editing and improvement of clarity. The author takes full responsibility for all scientific content, analyses, and conclusions.

Unterschrift:

_____

Tübingen, der

_____

Acknowledgments

I would like to thank everyone who contributed to the successful completion of this doctoral thesis. Primarily, I would like to thank my thesis supervisor Prof. Dr. Andreas Nieder for guiding me through my academic years, for introducing me to this exciting project, for all his help on theoretical issues and the valuable advice that helped me with the training of the monkeys. I have always enjoyed the constructive discussions that contributed to the success of this thesis. I would also like to acknowledge Prof. Dr. Uwe Ilg as the second reader of this thesis.

Furthermore, I would like to thank my family and friends for all their personal support throughout the last years.

I dedicate this thesis to my son, Darion Machts, and my wife, Antonia Machts.

Thank you,

Tobias Machts

Table of Contents

List of Abbreviations

AIP	**a**nterior **i**ntraparietal **a**rea
ANS	**a**pproximate **n**umber **s**ystem
AUC	**a**rea **u**nder the **c**urve
CCA	**c**anonical **c**orrelation **a**nalysis
CIP	**c**audal **i**ntraparietal **a**rea
dlPFC	**d**orso**l**ateral **p**re**f**rontal **c**ortex
DMC	**d**elayed **m**atch-to-**c**ategory
DMS	**d**elayed **m**atch-to-**s**ample
IPS	**i**ntra**p**arietal **s**ulcus
IT	**i**nferior **t**emporal cortex
LFP	**l**ocal **f**ield **p**otential
LIP	**l**ateral **i**ntraparietal **a**rea
MIP	**m**edial **i**ntra**p**arietal area
MUA	**m**ulti-**u**nit **a**ctivity
OFS	**o**bject **f**ile **s**ystem
PFC	**p**re**f**rontal **c**ortex
PPC	**p**osterior **p**arietal **c**ortex
RT	**r**eaction **t**ime
VIP	**v**entral **i**ntra**p**arietal area
vlPFC	**v**entro**l**ateral **p**re**f**rontal **c**ortex

Summary

This dissertation investigates how the brain of non-human primates encodes, retains and recognizes numerical information, combining behavioral and neuronal approaches. Understanding how numerical cognition emerges from distributed neural processing offers a unique opportunity to explore the interaction between working memory, recognition memory and decision-making – three core components of higher cognition.

Electrophysiological recordings from the dorsolateral prefrontal (dlPFC) and posterior parietal (PPC) cortices of rhesus monkeys performing a sequential delayed match-to-numerosity task revealed that single neurons encode recognition memory for numerical quantities. During comparisons of the sample stimulus with up to three test stimuli, the neurons exhibited distinct yet interacting patterns of familiarity modulation, including repetition suppression and match enhancement. These results suggest that recognition memory encompasses abstract, non-symbolic representations, and that the ability to detect familiarity is distributed across prefrontal and parietal networks.

At the population level, the interaction between dlPFC and PPC was further characterized by analyzing simultaneously recorded neuronal activity using canonical correlation analysis (CCA). This approach revealed dynamic, direction-specific coupling between both regions that varied across task phases. During stimulus encoding, information flow was predominantly feedforward from parietal to prefrontal areas, whereas maintenance periods were marked by more balanced interactions. These results suggest that numerical recognition arises from the flexible coordination of the frontoparietal network as a whole, rather than from the isolated processing of a single brain region.

Extending these findings to a comparative perspective, behavioral performance in the same task was examined across species, revealing that rhesus monkeys and carrion crows show comparable accuracy and numerical effects, such as distance and size effects. Despite these similarities, the temporal dynamics of decision-making differed systematically between species. Monkeys exhibited increasing reaction times across successive test phases, whereas crows responded progressively faster. Machine-learning analyses further demonstrated that reaction times carried information about

numerosity and task phase in monkeys but not in crows, suggesting species-specific strategies in numerical recognition and decision control.

Together, these studies provide a comprehensive, multi-level account of numerical cognition, covering from single-neuron encoding and network dynamics to behavior across species. By integrating neurophysiological and behavioral evidence, this thesis sheds light on abstract cognitive functions, such as number recognition.

Zusammenfassung

In dieser Dissertation wird erforscht, wie das Gehirn nichtmenschlicher Primaten numerische Informationen kodiert, speichert und wiedererkennt. Dazu werden neurophysiologische und verhaltensbezogene Ansätze kombiniert. Ein besseres Verständnis davon, wie numerische Kognition aus verteilter neuronaler Verarbeitung hervorgeht, eröffnet einzigartige Möglichkeiten, das Zusammenspiel von Arbeitsgedächtnis, Wiedererkennungsgedächtnis und Entscheidungsprozessen zu erforschen. Diese drei Komponenten sind von zentraler Bedeutung für höhere Kognition.

Elektrophysiologische Ableitungen aus dem dorsolateralen präfrontalen Kortex (dlPFC) und dem posterioren Parietalkortex (PPC) von Rhesusaffen während einer sequenziellen Delayed-Match-to-Numerosity-Aufgabe zeigten, dass einzelne Neurone an der Kodierung des Wiedererkennungsgedächtnisses für numerische Mengen beteiligt sind. Beim Vergleich eines Beispielreizes mit bis zu drei Testreizen wiesen Neurone in beiden Arealen charakteristische Aktivitätsmuster auf, die systematisch von der Vertrautheit der Reize moduliert wurden. Dazu gehörten sowohl „Repetition Suppression“ als auch „Match Enhancement“. Diese Ergebnisse legen nahe, dass das Wiedererkennungsgedächtnis nicht auf visuelle oder objektbezogene Reize beschränkt ist, sondern auch abstrakte, nicht-symbolische Zahlenrepräsentationen umfasst. Darüber hinaus deutet dies darauf hin, dass diese Funktion durch verteilte frontoparietale Netzwerke realisiert wird.

Auf der Populationsebene wurde die Interaktion zwischen dlPFC und PPC durch eine kanonische Korrelationsanalyse (CCA) simultan aufgezeichneter neuronaler Aktivität weiter charakterisiert. Dabei zeigte sich eine richtungsspezifische Kopplung, die sich über die verschiedenen Phasen der Aufgabe hinweg dynamisch veränderte. Während der Kodierungsphase dominierte ein feedforward gerichteter Informationsfluss vom PPC zum dlPFC, wohingegen in der Arbeitsgedächtnisphase eine ausgewogenere Kopplung zwischen beiden Arealen beobachtet wurde. Diese Befunde legen nahe, dass numerische Wiedererkennung nicht durch isolierte Hirnregionen realisiert wird, sondern auf flexiblen frontoparietalen Netzwerkdynamiken beruht.

In einer vergleichenden Perspektive wurde die Verhaltensleistung in derselben Aufgabe über verschiedene Spezies hinweg untersucht. Dabei zeigte sich, dass Rhesusaffen und Rabenkrähen eine vergleichbare Genauigkeit sowie klassische numerische Effekte (Distanz- und Größeneffekte) aufweisen. Trotz dieser Gemeinsamkeiten unterschieden sich die zeitlichen Dynamiken der Entscheidungsprozesse systematisch zwischen den Spezies. Während Rhesusaffen über aufeinanderfolgende Testphasen hinweg zunehmend längere Reaktionszeiten zeigten, reagierten Rabenkrähen zunehmend schneller. Analysen mittels maschinellen Lernens ergaben zudem, dass die Reaktionszeiten der Affen Informationen über die Aufgabenstruktur, insbesondere über die Anzahl der Testphasen, enthalten, während dies bei den Krähen nicht der Fall war. Dies deutet auf spezies-spezifische Strategien bei der Integration von Gedächtnis- und Entscheidungsprozessen hin.

Diese drei Studien liefern gemeinsam ein mehrstufiges Verständnis der numerischen Kognition: von der Einzelzellaktivität über die interareale Kommunikation bis hin zum Verhalten und dem Vergleich über Arten hinweg. Durch die Integration von neurophysiologischen und Verhaltensdaten leistet diese Arbeit einen Beitrag zum Verständnis abstrakter kognitiver Funktionen bei der Zahlenerkennung.

List of Publications

published

Machts, T., Grüb, J., & Nieder, A. (2025). Neuronal encoding of recognition memory for numerical quantities in macaque intraparietal and prefrontal cortices. *Progress in Neurobiology*, *251*, 102794.

Machts, T., & Nieder, A. (2026). Coordinated parieto-frontal neuronal communication is critical for abstract quantity judgments in primates. *Cell Reports*, *45*(4).

Machts, T., Kirschhock, M. E., & Nieder, A. (2026). Reaction-time signatures reveal divergent cognitive strategies underlying numerical decisions in monkeys and crows. *Cognition*, *273*, 106527.

Declaration according to § 5 Abs. 2 No. 8 of the Doctoral Regulations of the Faculty of Science

Contributions to Collaborative Publications

Last Name, First Name: _____

List of Publications

1. Machts, T., Grüb, J., & Nieder, A. (2025). Neuronal encoding of recognition memory for numerical quantities in macaque intraparietal and prefrontal cortices. *Progress in Neurobiology*, *251*, 102794.
2. Machts, T., & Nieder, A. (2026). Coordinated parieto-frontal neuronal communication is critical for abstract quantity judgments in primates. *Cell Reports*, *45*(4).
3. Machts, T., Kirschhock, M. E., & Nieder, A. (2026). Reaction-time signatures reveal divergent cognitive strategies underlying numerical decisions in monkeys and crows. *Cognition*, *273*, 106527.

Nr.	Accepted manuscript (peer-reviewed) yes/no	№ of authors	Position of candidate in list of authors	Scientific ideas by the candidate (%)	Data generation or/and programming by the candidate (if applicable) (%)	Analysis/ interpretation/ technical work by the candidate (if applicable) (%)	Paper writing done by the candidate (%)
1	yes	3	1	50	80	70	60
2	yes	2	1	60	100	80	70
3	yes	3	1 (shared first authorship)	50	50	70	60

This table can/should be supplemented with suitable continuous text if required.

I confirm that the above-stated is correct.

Date, Signature of the candidate

I/We certify that the above-stated is correct.

Date, Signature of the doctoral committee or at least of one of the supervisors

Part I.
Synopsis

1. Introduction

Our environment holds an abundance of sensory information. Therefore, finding a certain visual target among numerous irrelevant objects can be challenging. Cognitive control is required to mentally store a target and to select it from the surrounding stimuli by using its distinctive features, such as color, orientation, or shape (Chelazzi et al., 1993). Usually, we first evaluate the supposed target, mentally compare it with the features of the actual target representation, and finally decide whether it matches or not (Romo & Salinas, 2001). Theoretically, this comparison process must be carried out sequentially until the correct target is found – that is, each new visual input must be compared step by step with the stored target information that is persistently maintained in our short-term memory.

Both processes – holding the target in mind and deciding whether it is correct – are essential components of such visual search behavior. These major cognitive operations are known as working memory and decision-making, respectively. Working memory allows us to maintain and manipulate goal-relevant information in the absence of sensory input, whereas decision-making evaluates incoming stimuli and guides appropriate responses based on stored representations. When a stimulus is encountered that has been experienced before, both mechanisms interact to give rise to recognition memory, a process that enables us to identify familiar information by comparing new sensory input with previously stored representations. Recognition memory therefore bridges the functions of working memory and decision-making and allows efficient and adaptive behavior in dynamic environments.

Understanding how these cognitive processes are implemented at the neuronal level is essential for revealing the mechanisms that enable flexible behavior. Among them, working memory represents the core capacity that supports all higher cognitive functions, including recognition and decision-making.

1.1 Working memory

Working memory refers to the ability to temporarily store and manipulate information that is no longer present in the sensory environment (Baddeley, 1992). It provides the mental workspace for comparing, updating, and guiding behavior based on internal representations, thereby forming the foundation for complex cognitive operations such as reasoning, planning, and decision-making (Shettleworth, 2010).

A classical paradigm to study working memory in animals is the delayed match-to-sample (DMS) task. In this task, subjects are required to memorize a sample stimulus over a brief delay period until a test stimulus appears, and then indicate whether the test matches the sample. Because the sensory information is absent during the delay, successful performance depends on maintaining an internal representation of the sample.

Many factors can influence the behavioral performance in DMS tasks, like the duration of stimulus presentation (Grant, 1976), the duration between subsequent trials (Maki et al., 1977), the duration of the delay period (Miles, 1971) or the presentation of distractors (Jarvik et al., 1969). Usually, performance decreases with shorter stimulus duration and a higher delay phase as well as high similarity between sample and distractor stimulus. These effects illustrate that the capacity of working memory is limited and highly sensitive to distraction (Vogel et al., 2005).

Distractors – defined as task-irrelevant stimuli presented during the delay – are a useful tool to probe this limitation. Their disruptive effect depends on where, how long and with what kind of properties the distractor is presented to the subject. One common method is to present the distractor during the memory delay with a fixed or variable window length (Jacob & Nieder, 2014; Maki et al., 1977; Sakai et al., 2002). For example, Maki et al. (1977) demonstrated that brief illumination during the delay period reduced color matching performance in pigeons. Additionally, Jarvik et al. (1969) showed that in rhesus monkeys the distractor properties affect task performance. While positive interfering stimuli caused a slight increase in discrimination performance, neutral and negative interfering stimuli reduced task performance. Suzuki and Gottlieb (2013) confronted rhesus monkeys with a modified version of an oculomotor memory task, in which the monkeys had to remember the location of a peripheral target. The performance of the monkeys was unaffected if the distractor appeared within 900 ms after target onset in a delay period. In this situation, it played no role if the distractor appeared at near or far locations compared to the target. However, increasingly more errors were made if distractors became more similar in time and space to the target. Katsuki and Constantinidis (2012a) trained rhesus monkeys to perform a DMS task that required them to detect a salient stimulus in a visual display. In the study the distractor color of the visual display was systematically varied such as the performance of the monkeys was decreased when color similarity between salient stimulus and distractor

increases. Such findings highlight that working memory is not a passive buffer but an active process involving attentional control and filtering of irrelevant information.

Beyond general stimulus processing, working memory also plays a key role in more abstract domains such as numerical cognition. In sequential matching tasks, for example, animals must maintain a target numerosity across delay periods while comparing it to subsequently presented sample quantities (Nieder et al., 2002; Ditz & Nieder, 2016). The ability to preserve this internal representation despite interfering nonmatching stimuli is essential for accurate numerical judgments. Hence, numerical working memory provides a direct experimental window into the neural mechanisms of cognitive stability and flexibility.

1.2 Decision-making

While working memory enables the temporary maintenance of task-relevant information, effective behavior additionally requires mechanisms that evaluate this stored information and guide appropriate responses. This evaluative process – known as decision-making – determines how incoming sensory evidence is compared against internal representations and how a final choice is formed (Romo & Salinas, 2001). In this sense, decision-making is closely linked to working memory, because both processes depend on the integration of information across time and on maintaining relevant representations in the absence of sensory input.

The ability to make decisions is essential for goal-directed behavior. It enables us to pursue our goals in front of alternatives that are difficult to distinguish. According to Schall (2001) decisions are effortful, take time, require attention and deliberation, and are error prone. Furthermore, decision-making can be described as a sequential process. At the beginning sensory information must be encoded. After that the relevant features must be extracted and kept in mind. Finally, this new information must be compared to the previously stored information (Romo & Salinas, 2001). Depending on the result of the decision a communication with motor apparatus follows.

A common way to investigate decision-making in animals and humans is through visual search paradigms (Leon & Shadlen, 1998). In this task design subjects must distinguish a target stimulus from a variable array of distractor stimuli. Different approaches exist for how subjects report the target (Müller & Krummenacher, 2006):

1. The target is either present or absent, and the observer must make a target-present versus target absent decision as rapidly and accurately as possible.
2. The search display is presented just for a limited duration, and the observer must detect the target as good as possible.

The measured reaction time (RT) for these decisions can be used to create search RT functions when plotting against the display size (Treisman & Gelade, 1980). Such functions reflect the search rate, i.e. the search time per display item.

Another method to investigate decisions can be studied using the previously described DMS tasks. These tasks require the comparison of current sensory information with a stored template and the decision whether both correspond. Such sequential comparison processes can be conceptualized as match-nonmatch decisions and are well suited to study the interaction between working memory and decision-making mechanisms (Romo & Salinas, 2001). For example, Zhou et al. (2021) used dot motions of six different directions, which were separated by a learned category boundary. Analogues to other DMS tasks, the trained rhesus monkeys needed to respond when the learned categories of sample and test stimuli matched, or wait for second test stimulus when they did not match. This task design is also called delayed match-to-category (DMC) task.

As with working memory, our decision to identify the correct stimulus can be affected by distractors, which is reflected in search performance (e.g. RT). In visual search paradigms, visual search is more efficient when the target is fundamentally different from the other irrelevant visual objects. For example search is easy, when the subject has to look for a red circle among green circles (differ in color), a cross among a lot of triangles (differ in shape) or a vertical line among a lot of horizontal lines (differ in orientation) (Chan & Hayward, 2013). In this case the RT is independent of the total number of items and it feels like the target ‘pops out’ from other items (Chan & Hayward, 2013). However, our search performance becomes less efficient if the distractor shares more similarities with the target. For example, the probability to find a 45° oriented bar among a lot of 90° oriented bars is much higher, compared to find it among a set of 46° or 47° oriented bars (Verghese, 2001). In this case the RT increases with a higher display size and it feels like we have to sweep our attention around the stimulus display in order to find the correct target (Chan & Hayward, 2013). It becomes even more difficult when the target differs from the other objects not just by one feature (like the

orientation), but by a combination of features (conjunctive search) (Chan & Hayward, 2013).

DMC tasks that investigated sequential decisions came to the same result and found higher error rates with increasing similarity between matching and nonmatching stimuli (e.g. Freedman & Assad, 2006; Freedman et al., 2001; Zhou et al., 2021). Freedman et al. (2001) trained monkeys to categorize computer-generated stimuli as “cats” and “dogs”. By using a morphing system that systematically varied the stimulus shape, a precise definition of the category boundary was possible. The monkeys classified dog-like cats (60:40 cat:dog) or cat-like dogs (60:40 dog:cat) more error-prone than pure dog or pure cat morphs, i.e. the closer the stimulus was around the category boundary, the higher was also the error rate that the animals did. Similar results were also found for different motion directions (Freedman & Assad, 2006).

In the context of numerical cognition, decision processes translate into judgments about whether two numerical quantities are the same or different. These numerical match-nonmatch decisions require the animal to retain the target numerosity in working memory, compare it with each subsequently presented quantity, and finally decide if it corresponds to the remembered sample. Hence, numerical decisions provide a powerful model for studying how cognitive control, memory, and perceptual evaluation jointly contribute to higher-order cognition (Merten & Nieder, 2013).

1.3 Recognition memory

Recognition memory refers to the ability to identify whether a currently perceived stimulus has been encountered before. It thus represents a cognitive function that bridges working memory and decision-making: it requires the retrieval and evaluation of stored information to guide an explicit choice (Winters et al., 2008). While working memory maintains transient representations, recognition memory determines the familiarity or novelty of incoming sensory input based on these stored representations. Traditionally, recognition memory has been investigated using arbitrary visual objects, such as pictograms or images, where neuronal activity reflects stimulus familiarity or repetition (Brown & Aggleton, 2001; Rainer & Miller, 2000). For example, Miller et al. (1996) trained rhesus monkeys on a DMS task with a variable sequence of intervening visual objects. They were required to maintain the initial sample object in mind and to respond by a bar release whenever it was repeated. Other irrelevant objects that were

presented in between had to be behaviorally ignored by holding the bar. Importantly, the monkeys never knew if and when the matching stimulus might appear. While the monkeys were performing, the authors found prefrontal neurons that respond differentially to the same stimulus depending on whether it matched a previous stimulus or not. Therefore, this activity is informative about the prior stimulus. Other studies came to the same conclusion and found in DMS task prefrontal neurons that show different response patterns to a stimulus if it appears as a match or a nonmatch (Kusunoki et al., 2009; Pasternak & Zaksas, 2003; Qi et al., 2012). Therefore, these neural signatures offer an insight into how the brain distinguishes between novel and familiar information.

In the context of numerical cognition, recognition memory operates at an abstract, categorical level: the animal must recognize not a specific visual pattern or object, but a numerical quantity as familiar or repeated. This process relies on the integrity of the stored numerical representation in working memory and its comparison with the currently perceived numerosity. Hence, numerical recognition memory offers a unique opportunity to examine how the frontoparietal network supports the recognition of abstract quantities and how the brain encodes the familiarity of numbers independently of their physical appearance.

To understand how these cognitive functions are implemented in the brain, it is essential to identify their neural substrates. In primates, working memory, decision-making, and recognition memory all depend on coordinated activity within higher association cortices – most prominently, the prefrontal and posterior parietal cortices.

1.4 Neural correlates of working memory, recognition and decision-making

Complex cognitive operations such as working memory, recognition, and decision-making emerge from the coordinated activity of distributed cortical networks. Among these, the prefrontal cortex (PFC) and the posterior parietal cortex (PPC) play a central role. Both regions are part of the primate association cortex and exhibit overlapping neuronal activity patterns that reflect the temporary storage, evaluation, and comparison of information across time (Meister et al., 2013; Fuster, 2008).

The primate visual system is broadly divided into two processing streams: the ventral “what” pathway and the dorsal “where” pathway (Mishkin & Ungerleider, 1982). While the ventral stream, projecting into the inferior temporal cortex, supports object

recognition and non-spatial memory, the dorsal stream projects into the PPC and is specialized for processing spatial and quantitative information. The PFC, positioned at the top of this cortical hierarchy, integrates inputs from both pathways and interacts reciprocally with parietal regions to enable flexible cognitive control (Goldman-Rakic, 1988).

Neural correlates of working memory are classically characterized by persistent neuronal activity during delay periods in the absence of sensory stimulation (Goldman-Rakic, 1991). Such activity reflects the active maintenance of task-relevant representations and is modulated by distractors, attentional demands, and behavioral outcomes. In contrast, recognition memory involves changes in neuronal firing when a stimulus is re-encountered, typically manifesting as “repetition suppression” (Miller & Desimone, 1994; Riches et al., 1991; Xiang & Brown, 1998), but also other mechanisms for signaling familiarity have been observed, such as more selective tuning (Rainer & Miller, 2000; Freedman et al., 2006; Anderson et al., 2008), tuning peak enhancement (Woloszyn & Sheinberg, 2012; Lim et al., 2015), or even increases in firing rate (Tamura et al., 2017). These patterns reveal how the brain distinguishes familiar from novel information. Finally, decision-making is linked to time-dependent neural signals such as ramping activity, reflecting the accumulation of sensory evidence toward a categorical choice (Shadlen & Newsome, 1996; Gold & Shadlen, 2007).

Importantly, these three processes – maintenance, recognition, and decision – are not isolated functions but interacting components of a common cognitive architecture. The PFC and PPC exhibit both shared and distinct contributions to this architecture: while PFC neurons are crucial for representing abstract rules, maintaining information, and exerting top-down control, PPC neurons provide precise sensory and quantitative representations and contribute to evidence accumulation. Their reciprocal communication forms the frontoparietal network, a dynamic system supporting cognitive flexibility and goal-directed behavior (Fuster, 2008).

In the following sections, the specific roles of the prefrontal and parietal cortices will be described in greater detail, highlighting their neuronal signatures during working, recognition, and decision-related processing, and their interaction in guiding numerical judgments.

1.4.1 The prefrontal cortex

The PFC is an interconnected set of neocortical areas that are linked with most motor and sensory systems (except primary sensory and motor cortices), as well as a wide range of subcortical structures (Goldman-Rakic, 1988). It comprises several anatomically and functionally distinct subdivisions, most prominently the lateral PFC, the anterior cingulate cortex, and the orbitofrontal cortex.

Within this broader framework, the lateral PFC lies at the apex of the cortical hierarchy, underlies executive functions, and plays a key role in representing and organizing information for goal-directed behavior (Fuster, 2008). Furthermore, the lateral PFC is critically involved in tasks requiring binary choices and decisions that depend on the integration of multiple sources of information (Miller & Cohen, 2001; Fuster, 2008).

Due to distinct architectonic differences, the lateral PFC can be further divided into ventrolateral (vlPFC) and dorsolateral (dlPFC) regions (Tanji & Hoshi, 2008). The vlPFC is associated with response control (Mishkin et al., 1969; Rushworth et al., 1997), whereas the dlPFC integrates spatial and temporal information and plays a key role in working memory (Fuster, 2008). Consequently, the lateral PFC is considered a key candidate for exerting top-down control over lower-level visual areas such as the PPC (Anderson & Green, 2001).

Unless stated otherwise, the term “PFC” refers to the dlPFC throughout the result and discussion part of this synopsis.

1.4.2 Representation of working memory in prefrontal cortex

In delayed response tasks, the PFC shows a high proportion of neurons presenting stimulus selective persistent spiking activity during the delay period (Goldman-Rakic, 1991). This sustained activity is said to be a physical instance of our actual working memory, i.e. it reflects the sustained representation of working memory content. Working memory related activity in the PFC has been found for almost all modalities across time and space. To name a few: for objects and visual pictures (Miller et al., 1996), for spatial information (Goldman-Rakic, 1995), for motions (Zaksas & Pasternak, 2006) and even highly abstract categories like numerical quantities (Nieder et al., 2002). Thus, the PFC is involved in all working memory tasks irrespective of the task specifics or properties (Fuster, 2008). More importantly, on the basis of the persistent spiking activity it is possible to predict the behavioral, task-related outcome. The persistent

activity is in error trials compared to correct trials severely reduced. Thus, the working memory content holds not just stimulus related information, but also information about goal-directed actions. This close link between neuronal activity that underlies the working memory and the resulting behavior is unique for the PFC and not found for any other brain region (Fuster, 2008).

Additionally, the important role of the PFC for goal-directed behavior is several times shown by lesion studies: Both, permanent and reversible frontal damage lead to a loss to maintain goal-relevant information in rhesus monkeys and humans (Fuster, 2008; Milner, 1963). Especially lesions at the principal sulcus in the lateral PFC result in a severely impaired working memory (Castner et al., 2004). Important signs are longer learning intervals and impaired performance with increasing delay intervals (Fuster, 2008). However, the above described effects were strongly reduced when the task required no manipulation of task information or no monitoring of task progression (Dash et al., 2007). Thus, the PFC acts more as a operator for higher cognitive control functions such as organization and manipulation of working memory or attentional filtering (Lara & Wallis, 2015; Lennert et al., 2011).

A further important indication of PFC damage is the increased distractibility (Duncan et al., 1996), i.e. subjects seem to be unable to focus their attention on a certain target, when another competing object or event is presented. A variety of neurophysiological studies showed that the intact PFC has the ability to maintain information over distracting events which is reflected in a persisting coding of the initial representation and reduced response to the distractor (di Pellegrino & Wise, 1993; Everling et al., 2002; Lennert & Martinez-Trujillo, 2011; Miller et al., 1996; Qi et al., 2010; Suzuki and Gottlieb, 2013). For example, Suzuki and Gottlieb (2013) confronted rhesus monkey with a memory saccade task in which a salient distractor was flashed at a variable timing and location during the memory delay. PFC neurons barely responded to the distractor and represented constantly the sample stimulus. Furthermore, the neuronal activity was closely correlated with the behavioral performance.

In summary, the PFC has the ability to resist against distractor information, which supports our working memory robustness (Murray et al., 2017) and finally our filtering performance (Everling et al., 2002).

1.4.3 Representation of decision-making in prefrontal cortex

Decision-making is closely linked with our working memory (and vice versa). For both processes, we must integrate information over time. Therefore, it can be challenging to distinguish neuronal activity related to working memory from those related to decision-making. For example, persistent neural activity during sensory stimulus absence (delay period) was interpreted by Shadlen and Newsome (1996) as the time point in which the decision was made and kept in memory until choice report. This mechanism is similar to the working memory. Hence, it was postulated that these two processes might be shared by the same cortical mechanisms as well as anatomical structures (Wang, 2008).

Although it seems to be difficult to assign neural activity to working memory or decision-making, there are differences between both cognitive processes. While working memory is associated with sustained activity that carries integrated information during stimulus absence (see chapter above), decision-making is typically related to ramping neural activity (Wang, 2008). This ramping activity is reflected by a gradual accumulation of the respective perceptual evidence while the sensory information is still present (Wang, 2008). The magnitude of the accumulated evidence forms distributions for the target and non-targets, whereas the target selection is dedicated to the larger of the alternative accumulated values (Gold & Shadlen, 2007; Schall, 2001). For example, Gold and Shadlen (2000) found ramping activity in neurons of the frontal eye field during motor preparation in a two-alternative visual motion discrimination task. Recent studies of single-neuron recordings in the rodent PFC and of biophysical models with frontoparietal circuits reveal that the representation of accumulated evidence is more categorical than graduated (Hanks et al., 2015; Murray et al., 2017), i.e. the ramping activity in PFC is marked by a characteristic steep slope. In match-nonmatch decisions tasks prefrontal neurons act also as a candidate source of decision signals (Zhou et al., 2021). These neurons show a nonlinear integration of currently visible and remembered stimuli, which is highly correlated with the behavioral outcome (Zhou et al., 2021). Freedman and Assad (2011) were the first who postulated that decision neurons could rather reflect a perceptual categorization than a motor preparation (in order to report the decision). Merten and Nieder (2012, 2013) confirmed this by separating the decision period from the phase which is intended for motor preparation. Thus, decisions can

either be made in an action-based or report-independent framework, and potentially processed like abstract categories (Merten & Nieder, 2012).

The importance of the PFC for decisions is also reflected in lesion studies: Humans seem to have severe impairments in real-life decision-making which is characterized by repeated engagement in high-risk behaviors (known as acquired sociopathy, Manes et al., 2002). In monkeys led a bilateral excitotoxic PFC lesion to a reduced probability to choose the high-reward option in a three-armed bandit task, compared to a control group with an intact PFC (Rudebeck et al., 2017). Depending on different aspects of decision-making, the PFC can separated into three regions (Krawczyk, 2002): The orbitofrontal and ventromedial areas of the PFC are most relevant for decisions based on reward values (‘Which is the best option?’). The dorsolateral part is important for decisions when several sources of information is needed. The anterior and ventral cingulate cortex is relevant in conflicting options, as well as signaling outcome-relevant information. In summary, the PFC plays a key role in complex decision-making, which is reflected in characteristic neural activity (Murray et al., 2017).

1.4.4 Representation of recognition memory in prefrontal cortex

Beyond its role in working and decision processes, the PFC also contributes critically to recognition memory. Neurons in lateral and ventrolateral prefrontal areas differentiate between familiar and novel stimuli, showing repetition suppression or match enhancement depending on task context (Rainer & Miller, 2000; Miller et al., 2006). Match enhancement is the increased neuronal activity that occurs when a stimulus matches a previously encountered stimulus stored in memory. This active mechanism engages in favor of sought-after stimuli, ensuring that repetitive, irrelevant nonmatch stimuli are ignored (Engel & Wang, 2011). By contrast, repetition suppression refers to a decrease in response magnitude with repetition of a familiar stimulus (Miller et al., 1996). These activity patterns reveal that the PFC evaluates the familiarity of incoming stimuli and integrates such evaluations into behavioral responses.

While sensory recognition memory has been studied using arbitrary pictogram stimuli (Brown & Aggleton, 2001), the memory necessary for recognizing abstract categories, such as numerical quantities, in a sequence of numerosities has not been investigated. In contrast to arbitrary visual stimuli, the special feature of numerical quantities is that they are semantically related to each other by ordinal magnitude, also known as the

“mental number line” (Nieder, 2020; see chapters below). Recognition memory of numerical quantities therefore depends heavily on the representation of target numerosities as opposed to neighboring non-target numerosities, which is a unique feature of parameterized quantities and a central topic of this thesis.

Together, the PFC forms an essential node that transforms maintained information into evaluative judgements. They bridge working memory, recognition and decision-making functions within a single cortical system.

1.4.5 The posterior parietal cortex

The PPC, together with the inferior temporal cortex (IT) and PFC, is one of the three major associative regions in the brain and has several higher-order functions. It integrates signals from subcortical areas, combines input from somatosensory, auditory, visual and motor cortices (Whitlock, 2017) and is tightly interconnected with areas of the PFC (Cavada & Goldman-Rakic, 1989). The PPC is separated by the intraparietal sulcus (IPS) into the superior and inferior partial lobe. The IPS is based on cytoarchitectural and functional differences divided into medial (MIP), lateral (LIP), ventral (VIP), caudal (CIP) and anterior (AIP) areas (Grefkes & Fink, 2005). The PPC can encode visual features such as direction (Fanini & Assad, 2009), color (Toth & Assad, 2002), form (Fitzgerald et al., 2011) and mediate abstract cognitive capacities, including the representation of numerical quantity (Nieder et al., 2002). The PPC is also involved in further cognitive processes, including sensorimotor integration, spatial attention, spatial navigation and early motor planning, as well as working memory, recognition memory and decision-making (Whitlock, 2017; Rawley & Constantinidis, 2010).

1.4.6 Representation of working memory in the posterior parietal cortex

Similar to the PFC, the PPC contains a high proportion of neurons that show persistent activity in memory delay tasks during stimulus absence (Constantinidis & Steinmetz, 1996; Rawley & Constantinidis, 2009). For example, several studies showed that LIP neurons represent the sample stimulus throughout spatial working memory tasks (Chafee & Goldman-Rakic, 1998). VIP neurons selectively respond to numerical quantity and carry this information over the delay period (Nieder et al., 2002; Nieder & Merten, 2007; Nieder & Miller, 2003, 2004; Okuyama et al., 2015; Tudusciuc & Nieder, 2009). Like the PFC, the neuronal activity, associated with working memory, can be

classified as “sustained” (Quintana & Fuster, 1992). Thus, the representation of the neural correlates of working memory is not exclusive to the PFC, and a comparison of studies revealed a high degree of similarity between the two brain areas. For example, there are similar percentages of activated neurons, similar response magnitudes, similar temporal responses, and similar tuning for spatial objects (Chafee & Goldman-Rakic, 1998). However, the PFC seems to be more behaviorally relevant than the PPC, inasmuch as cooling of the PFC led to stronger performance decreases in spatial working-memory tasks compared to cooling of the PPC (Chafee & Goldman-Rakic, 2000).

1.4.7 Representation of decision-making in the posterior parietal cortex

Neuronal responses of the PPC are also known to represent correlates of decision-making (in this case area LIP, Shadlen & Newsome, 1996; Yang & Shadlen, 2007). Similar to the PFC, these neurons accumulate sensory information for forming a decision, proven for many useful quantities like value, benefit or reward expectation (Dorris & Glimcher, 2004; Platt & Glimcher, 1999; Sugrue et al., 2004). However, the representation of the accumulated evidence is graded and not categorical like in the PFC (Hanks et al., 2015; Murray et al., 2017). Further differences between PFC and PPC were described by Zhou et al. (2021) for sequential decisions in a DMC task. While neurons from area LIP and MIP were more involved in stimulus evaluation and motor planning, respectively, the PFC was prominently responsible for the match-nonmatch decision.

1.4.8 Representation of recognition memory in the posterior parietal cortex

Importantly, the PPC also contributes to recognition memory. PPC neurons exhibit a modulation by the match or nonmatch status of a remembered stimulus (Rawley & Constantinidis, 2010; Steinmetz et al., 1994). These responses may indicate the reactivation of stored sensory traces, allowing the PPC to participate in recognizing previously encountered information.

Although the PPC and PFC have a lot in common when it comes to representing recognition memory, differences in neuronal response have been identified when multiple stimuli are presented sequentially. While PFC neurons respond minimally to nonmatching items (Di Pellegrino & Wise, 1993; Lennert & Martinez-Trujillo, 2011; Qi

et al., 2010; Suzuki & Gottlieb, 2013), parietal cells represent the most recent stimulus, regardless of its relevance to the current task (Constantinidis & Steinmetz, 1996; Katsuki & Constantinidis, 2012b; Suzuki & Gottlieb, 2013). Therefore, the PPC encodes distractor information.

However, the distinct roles of the PFC and PPC appear to depend on the precise circumstances. To investigate the influence of distractors on the processing of numerical quantity, Jacob and Nieder (2014) modified the DMS task by inserting a behaviorally irrelevant number display during the delay period. The monkeys were able to ignore the distractor number displays most of the time, but they made more errors than when there were no distractors present. This is likely due to incomplete suppression of all distractor influences. During the task, single-unit activity was simultaneously recorded from the PFC and the VIP area. Throughout the first delay period, PFC neurons encoded the relevant sample number with sustained activity. However, when the distractor number was subsequently displayed, these neurons were unable to resist interference and began to respond more strongly to the distractor numbers. These findings suggest that the information was overwritten. Surprisingly, in the second delay after the distraction, the sample number representation (target) was regenerated. In contrast, neurons of the VIP area were nearly unaffected by the distractor and consequently maintained the working memory representation of the sample number. These findings are different to the previous introduced studies, which have argued for the PFC as the suppressor of interfering information.

1.5 Functional interaction of prefrontal and parietal cortices

The PFC and PPC are tightly reciprocally interconnected and show simultaneous activation during several cognitive functions, such as memory maintenance, recognition, and decision-making. Neurons in each area exhibit delay-period activity reflecting working memory, categorical selectivity for decision outcomes, and modulation by stimulus familiarity during recognition (see chapters above). Therefore, these two brain areas are often seen as a functional unit and called the frontoparietal network (Katsuki & Constantinidis, 2012b). Yet, despite these similarities, their contributions differ in temporal dynamics and functional emphasis. While the PFC contributes to the active maintenance of task-relevant information, rule representation, and top-down control, the PPC provides detailed sensory representations that serve as the foundation for these higher-order functions.

Crucially, effective cognition depends on the dynamic communication between these regions. Studies using simultaneous recordings have shown alternating patterns of feedforward and feedback interaction between PPC and PFC, reflecting shifts between sensory encoding and top-down control (Salazar et al., 2012; Buschman & Miller, 2007). For example, Salazar et al. (2012) recorded spiking activity and field potentials (LFP) from the PFC and PPC while monkeys are engaged in a spatial DMS task. Through the calculation of a coherence selectivity index it was possible to quantify how much mutual information of the target was between the prefrontal and parietal neurons. They found that there was an increase in mutual information of the memorized stimulus, especially during the delay period. Moreover, the information flow was primarily from the PPC to PFC (bottom-up), which is consistent with idea that the PPC (or other sensory cortices) acts as an information storage and the PFC has access to this information (Lara & Wallis, 2015). A further study confronted monkeys with a rule-based categorization task and correlated the corresponding spiking activity from PFC and PPC at different time lags (Crowe et al., 2013). The correlation was the strongest, when the PPC activity was later compared to the PFC activity, i.e. the information about the rule-dependent categories was selectively transmitted top-down from prefrontal to parietal neurons. A similar method to investigate brain interactions was described by Semedo et al. (2022). The authors correlated spiking activity of simultaneous recorded single cells from early and midlevel visual brain areas by a dimensionality reduction approach (canonical correlation analysis, CCA). By systematically correlating the activity at shifted time points, it was possible to identify feedback (top-down) and feedforward (bottom-up) dominated periods under stimulus control. The population interactions were particularly feedforward-dominated shortly after stimulus onset and feedback-dominated during spontaneous activity. Thus, the direction of the interaction depend on the cognitive process. For example, the processing of sensory information (e.g. stimulus onset) might be more dominated by a flow of information from PPC to PFC, while filtering (e.g. distractor information) might involve a flow of information from PFC to PPC.

The importance of prefrontal top-down control was highlighted by a recent biophysical model examining working memory and decision-making in a frontoparietal circuit model during distractor processing (Murray et al., 2017). They found out that the varying degree of PFC feedback (i.e. top-down signal from PFC to PPC) can either strengthen the presentation of the target (via feedback excitation) or can used to filter out the distractor information (via feedback inhibition). Combined stimulation-imaging in

humans during distractor presentation supports also the importance of PFC feedback to enhance target representation (Feredoes et al., 2011). Bidirectional interactions in the frontoparietal network were also found for sensorimotor processing (Siegel et al., 2015). In term of processing distractor numerical quantity, the contrary results of Jacob and Nieder (2014) also emphasize the interaction between PFC and PPC. In this case the PFC had represented more distractor information, whereas parietal cells did not. It was suggested that the activity of the PFC does not reflect working memory exclusively, but rather works as a top-down signal to influence other areas of the brain, like the PPC. This could mean that the PFC needs both representations of relevant and distractor information to filter (for example attentively, Lennert et al., 2011) and guide stimulus selection for the appropriate behavioral response.

In summary, the coordinated activity of PFC and PPC supports the neural mechanisms of working memory, recognition, and decision-making. Understanding how this network encodes and exchanges information about abstract quantities provides the conceptual foundation for the following chapter on numerical cognition.

1.6 Numerical cognition – processing of numerical quantities

The ability to perceive, represent and manipulate numerical information is a defining feature of advanced cognition. It allows organisms to estimate quantities, compare magnitudes and make adaptive decisions in both natural and artificial environments. From foraging and social interactions to symbolic mathematics, number-related abilities influence behavior in a variety of species and contexts (Dehaene et al., 1998; Nieder, 2016).

At the neuronal level, numerical cognition offers a distinctive model for investigating abstract representation and cognitive control. Unlike concrete sensory features such as color or shape, numerical magnitude is a conceptual property that must be extracted from varying visual or auditory input. This process involves the collaboration of parietal and prefrontal regions – the same frontoparietal network underlying working memory, recognition and decision-making processes. Within this network, neurons in the IPS and the dlPFC encode the numerical value of stimuli in a modality-independent and tuned manner, forming the basis of what has been termed “number sense” (Nieder & Miller, 2004; Roitman et al., 2007).

Importantly, numerical processing involves the encoding of quantity, as well as memory-based comparison and recognition. When a numerical quantity must be matched to a previously seen value, the task engages both working memory, which maintains the target numerosity, and decision-making and recognition, which evaluate current input relative to stored representations. Therefore, numerical cognition provides a framework that can be used to study the interaction of these fundamental cognitive functions at a neuronal level.

The following sections will examine the concept of numerical representation in more detail. They will outline behavioral and neuronal evidence for number selectivity in animals and humans, setting the base for the experimental studies presented in this cumulative thesis.

1.6.1 Foundations of numerical cognition

Numerical quantities, i.e. numerosity, is an abstract quantitative category and defined as the countable number of elements in a set. Within the animal kingdom many species show the ability to discriminate numerical quantity, including mammals (Cantlon & Brannon, 2006; Nieder et al., 2002), fishes (Agrillo & Bisazza, 2018), birds (Ditz & Nieder, 2016), reptiles (Gazzola et al., 2018; Lin et al., 2021), amphibians (Stancher et al., 2015) and invertebrates (Howard et al., 2018). A common method to investigate the numerical competence is the delayed match-to-numerosity task in which the animals are confronted with a specific number of items (set-sizes) in visual multiple-dot displays (Nieder et al., 2002). These dot displays are controlled for different low-level features, such as cumulated dot area and dot density, to prevent that non-numerical cues are used for task solving (Nieder et al., 2002). The behavioral representation of numerical quantities is then reflected by the numerosity judgement relative to nonmatching displays consisting of both smaller and larger numerosities (e.g. 2 and 4 for sample numerosity 3), resulting in bell-shaped behavioral performance curves. The averaged numerosity judgement of rhesus monkeys is almost perfect for the smallest numerosities but decreases towards larger numerosities (Nieder, 2018), i.e. it exists a systematic decrease in discriminability with increasing numerical size (Dehaene et al., 1998; Nieder, 2018; Viswanathan & Nieder, 2013). This effect is the so-called numerical size effect and lead to performance curves that become progressively wider for lager numerosity set-sizes. In rhesus monkeys, the discrimination threshold was found to be between 4 and 5 when using a criterion of 60 % correct in performance (Nieder et al.,

2002). Thus, the monkeys are able to discriminate numerical values 1 to 4 with a numerical distance of 1 between match and nonmatch, but failed for numerosities of 5 and higher. However, when the numerical distance between match and nonmatch display increases, the performance increases as well, i.e. the more distant two compared numerosities are, the better becomes the discrimination performance, an effect that is known as numerical distance effect.

Both the numerical distance effect and the previous described numerical size effect, are classical hallmarks of the approximate number system (ANS), which estimates numerosities with: 1.) an approximation, 2.) systematically less precision for increasing numerical values and 3.) no set-size limit (Cantlon and Brannon, 2006; Dehaene, 1992; Nieder, 2018). The ANS contrasts with an object file system (OFS), which provides precise number representations but fails to do so for quantities greater than four items. It has been reported for human children (Feigenson et al., 2004) and young chicks (Rugani et al., 2008). So far, monkeys show no signs of an OFS during absolute numerosity discrimination (Nieder, 2018). This is supported by the findings of Merten and Nieder (2009). They trained two rhesus monkeys to discriminate numerosities between 1 and 5. After exhibiting competent performance, the monkeys were confronted in transfer trails to distinguish novel and larger numerosities up to 30. Nevertheless, both animals were able to discriminate the new set of numerosities with similar judgment characteristics as for the smaller and trained numerosities. Most importantly, they showed no upper numerosity limit and no sudden change in performance between small numerosities and large numerosities, which would be expected for a transition from the OFS to the ANS. Besides trained monkeys (Beran, 2007; Cantlon & Brannon, 2006; Dehaene et al., 1998), the ANS is reported in infants (Xu et al., 2005), adult humans prevented from verbal counting (Dehaene et al., 1998) and in innumerate adult humans (Frank et al., 2008).

Altogether, the precision to discriminate numerosities increases with both smaller numerosity set-sizes and increasing numerical distance. Both effects result in bell-shaped behavioral performance curves that show an asymmetry, when plotted on a linear number scale. Consistent with the Weber-Fechner law (Fechner, 1860) a logarithmic scaling of the performance functions result in symmetric peak functions (Merten & Nieder, 2009; Nieder & Miller, 2003; Ramirez-Cardenas et al., 2016). This

logarithmic number representation is not exclusively observed in rhesus monkeys, but also in crows (Ditz & Nieder, 2016) and in pre-school children (Siegler & Booth, 2004).

Additionally, the numerical distance and numerical size influence not just the performance but also the corresponding reaction time: rhesus monkeys and from counting prevented humans response typically faster to smaller numerosities whereas larger quantities lead to a progressively increase in response latency together with a decrease in performance (Mandler & Shebo, 1982; Moyer & Landauer, 1967; Nieder et al., 2002).

1.6.2 Neuronal correlates of numerosity in frontoparietal networks

Ideal candidates for processing and storing numerosity are the PPC and dlPFC (Nieder et al., 2006, 2002; Nieder & Merten, 2007; Roitman et al., 2007; Sawamura et al., 2002), since both brain areas receive highly processed input from nearly all senses, which is necessary for abstractly representing numerosity across multiple sensory modalities. Furthermore, the dlPFC and PPC are anatomically (Barbas et al., 1985) and physiologically (Quintana et al., 1989) interlinked. Several studies showed that so-called number neurons in the monkeys' dlPFC and the VIP of the PPC selectively respond to numerosity and maintain the numerical information over temporal delays (e.g. Nieder et al., 2002; Nieder & Merten, 2007; Nieder & Miller, 2003, 2004; Okuyama et al., 2015; Tudusciuc & Nieder, 2009), even if it is numerosity zero (Ramirez-Cardenas et al., 2016). Interestingly, for the neuronal representation it plays no role in which spatiotemporal format (Nieder et al., 2006) or sensory modality (Nieder, 2012) the numerosity is presented. The proportions of neurons that selectively respond to numerosity range in the dlPFC between 20 - 30 % and in the VIP between 15 - 20 % (Nieder et al., 2002). Typically, the number neurons respond most strongly to their preferred numerosity and show a reduced discharge rate to adjacent numerosities resulting in a bell-shaped neural response function similar to the functions for the behavioral response (Merten & Nieder, 2009; Nieder & Miller, 2003, 2004).

Furthermore, the neuronal tuning functions widen with increasing numerical size (Nieder et al., 2002) and – in accordance with the Weber-Fechner law – are best described when plotted on a logarithmic number line (Nieder, 2016). Nieder (2016) hypothesized that this kind of logarithmic representations has two advantages:

1. The variability of the representations is constant across the number line, which ensures that the precision to discriminate numerosities is proportional to the number size. Thus, small numerosities are represented with high precision while large numerosities are more blurred.
2. A logarithmic compression enables even relatively few neurons to process increasingly large numerosities.

2. Research aims

This dissertation aims to elucidate how the primate frontoparietal network supports the active search and recognition of numerical quantities. While previous research has established that both the dlPFC and the VIP encode abstract numerical information, the precise division of labor between these areas, their reciprocal communication, and their relation to numerical behavior across species remain puzzling.

To address these open questions, this work combines three complementary studies, each focusing on a different level of neural and behavioral organization:

1. Single-neuron encoding of numerical recognition memory

   The first study ("*Neuronal encoding of recognition memory for numerical quantities in macaque intraparietal and prefrontal cortices"*) investigated how neurons in the VIP and dlPFC encode and maintain numerical information during a sequential delayed-match-to-numerosity task. By comparing tuning stability and match-nonmatch selectivity, the study identifies the neural correlates of recognition memory for numerosities, delineating their distinct contributions to processing in the parietal and frontal cortices.

2. Dynamic communication within the frontoparietal number network

   The second study ("*Coordinated parietal–frontal neuronal communication is critical for numerical judgments in primates"*) shifts focus from single-cell to population-level dynamics. Using simultaneous recordings from the VIP and dlPFC, and applying canonical correlation analysis (CCA), the study characterizes how neuronal interactions evolve during numerical working memory. The results of these analyses reveal a bidirectional coordination between the parietal and frontal areas, and demonstrate how this interaction breaks down during error trials.

3. Comparative behavioral signatures of numerical cognition

   The third study ("*Reaction-time signatures reveal divergent cognitive strategies underlying numerical decisions in monkeys and crows"*) places number discrimination in nonhuman primates in a comparative cognitive context, extending them to a cross-species behavioral comparison between rhesus monkeys and carrion crows performing the same delayed match-to-numerosity task. Both species demonstrated comparable levels of accuracy and exhibited features characteristic

of the approximate number system, including numerical distance and size effects. However, their reaction time patterns diverged systematically: the monkeys became slower across successive test phases, while the crows became faster. This suggests that the two species have different strategies for processing and retrieving numerical information.

Together, these studies establish a multi-level framework that links the neuronal encoding of numerical information, the communication between different brain regions, and the behavioral expression of numerical cognition. They demonstrate that effective numerical judgements depend not only on local representations within single cortical areas, but also on the coordinated interaction of these representations across the frontoparietal network. Ultimately, they show that these judgements depend on different cognitive strategies that exist across species.

3. Results

The following chapters present the main results of three manuscripts in sequence; progressing from neuronal coding (study 1, published), to network communication (study 2, published), and finally to cross-species behavior (study 3, published). The full publications are attached at the end of the thesis (see Individual studies).

3.1 Single-neuron encoding of numerical recognition memory

This study investigated how neurons in the macaque VIP and dlPFC encode and maintain numerical information during a sequential delayed-match-to-numerosity task. Both monkeys performed the task reliably and significantly above a chance level of 25 % (monkey 1: 27 sessions, 61.32 ± 1.36 %, $p < 0.001$; monkey 2: 33 sessions, 60.95 ± 0.58 %, $p < 0.001$, Binomial test), demonstrating that they were able to memorize a sample numerosity and to identify it within a sequence of up to three test displays. Behavioral performance showed characteristics of the approximate number system. Accuracy was highest for the smallest numerosity (numerosity 1: monkey 1, 95.02 ± 1.12 %; monkey 2, 96.01 ± 0.52 %) and declined systematically with increasing sample numerosity (numerosity 4: monkey 1, 56.31 ± 1.96 %; monkey 2, 64.57 ± 1.99 %). Errors occurred most frequently for nonmatching numerosities that were numerically adjacent to the sample numerosity, while more distant numerosities were more easily rejected, reflecting a clear numerical distance effect (Nieder & Miller, 2003; Merten & Nieder, 2009). In addition, increasing numbers of intervening nonmatch stimuli led to a gradual decrease in performance, indicating growing demands on recognition memory (two-way ANOVA with the factors 'number of nonmatches' and 'sample numerosity'; monkey 1: $F(3,431) = 60.1$, $p < 0.001$; monkey 2: $F(3,527) = 66.49$, $p < 0.001$). Nevertheless, performance in all trial types and for all sample numerosities remained significantly above chance (for all sample numerosities and trial types, $p < 0.05$, Binomial test), confirming that the monkeys successfully searched for and recognized the target numerosity even under increased cognitive load.

While performing the task, single-unit recordings were obtained from 372 neurons in the VIP and 509 neurons from the dlPFC. Visual response latencies were significantly shorter in VIP than in dlPFC, suggesting earlier sensory processing in the parietal area (median VIP: 111 ms, median PFC: 124 ms, Wilcoxon rank sum test, $Z = 2.6619$, $p = 0.0078$, two-sided). Neuronal activity was analyzed within time windows that excluded

potential motor-related signals. Within these windows, a subset of neurons in both areas showed clear numerosity selectivity during sample presentation (VIP: 34/254, 13.4 % of the recorded neurons; PFC: 68/369, 18.4 % of the recorded neurons). The proportion of numerosity-selective neurons was comparable to previous studies (Viswanathan and Nieder, 2015; Ramirez-Cardenas et al., 2016), and the tuning curves during the sample period were robust and stable in both regions, with no significant differences in tuning reliability between VIP and dlPFC (cross-validated tuning functions; Wilcoxon rank-sum test, $Z = -0.0674$, $p = 0.9462$, two-sided).

3.1.1 PFC neurons differentiate between relevant and irrelevant quantities, whereas VIP neurons do not

Pronounced differences between the two brain areas emerged during the test periods. In dlPFC, numerosity-selective neurons largely preserved their tuning when the matching numerosity was presented in the first test phase. Tuning curves in this match condition closely resembled those observed during sample presentation, both in terms of preferred numerosity and tuning width (sigma-values derived from fitted Gauss functions; median sample: $\sigma = 1.24$, median match: $\sigma = 1.45$, Wilcoxon signed-rank test, $Z = -1.8392$, $p = 0.1977$, two-sided, with Bonferroni correction, $n = 68$). In contrast, when a nonmatching numerosity was shown, PFC tuning deteriorated: tuning curves became broader and less selective (median match: $\sigma = 1.45$, median nonmatch: $\sigma = 2.01$, Wilcoxon signed-rank test, $Z = -2.4686$, $p = 0.0407$, two-sided, with Bonferroni correction, $n = 68$). This loss of selectivity was specific to nonmatch conditions. In VIP, however, numerosity tuning degraded from the sample to the test phase irrespective of stimulus relevance. Both match (median sample: $\sigma = 1.42$, median match: $\sigma = 2.35$, Wilcoxon signed-rank test, $Z = -2.9491$, $p = 0.0096$, two-sided, with Bonferroni correction, $n = 34$) and nonmatch (median sample: $\sigma = 1.42$, median nonmatch: $\sigma = 2.58$, Wilcoxon signed-rank test, $Z = -3.1372$, $p = 0.0051$, two-sided, with Bonferroni correction, $n = 34$) presentations led to flattened and broadened tuning curves, indicating a general decline in numerical selectivity.

These effects were quantified using a match-versus-nonmatch index based on firing rates to the preferred numerosity. In dlPFC, this index (median 0.26) was significantly larger than chance (median 0.17) ($Z = 0.9567$, $p = 0.0209$, Wilcoxon signed-rank test, $n = 68$), demonstrating a robust differentiation between target and irrelevant numerosities. In VIP, the index (median 0.18) did not differ from chance (0.15) ($Z =$

0.9935, p = 0.3387, Wilcoxon signed-rank test, n = 34), indicating an absence of systematic match-nonmatch discrimination at the single-neuron level.

Further analyses distinguished neurons showing match enhancement from those showing match suppression. Match enhancement, i.e. an increased response to repeated target numerosities, was observed exclusively in dlPFC neurons, whereas match suppression occurred in both dlPFC and VIP with comparable magnitudes. Thus, only dlPFC neurons actively enhanced responses to behaviorally relevant numerosities, while both areas exhibited suppression for repeated stimuli.

Population-level analyses supported these findings. Percent explained variance analyses across all recorded neurons revealed that both dlPFC and VIP populations represented match information during sample and delay periods. During the test phase, however, match information increased and remained dominant in PFC ($Z = 3.8263$, $p < 0.001$, Wilcoxon signed-rank test, n = 369), while VIP populations did not differentiate between match and nonmatch information ($Z = 1.6477$, $p = 0.0994$, Wilcoxon signed-rank test, n = 254). This suggests that only dlPFC populations encode the behavioral relevance of numerical stimuli.

Support vector machine classifiers trained on sample-period activity further highlighted this distinction. Numerosity decoding accuracy was higher in dlPFC (mean 60 ± 0.66 %) than in VIP (53 ± 0.66 %) during the sample period (permutation test, $p < 0.001$, α-corrected, n = 100) and remained significantly higher for matching test stimuli. In PFC, decoding accuracy was significantly better for match (42 ± 0.81 %) than for nonmatch conditions (35 ± 0.64 %) (permutation test, $p < 0.001$, α-corrected, n = 100), whereas VIP decoding performance did not differ between these conditions (match: 37 ± 0.64 %, nonmatch: 38 ± 0.60 %, permutation test, $p = 0.283$, α-corrected, n = 100). Moreover, classifier-derived tuning functions were narrower for match than for nonmatch conditions in dlPFC, an effect absent in VIP.

3.1.2 PFC neurons sustained tuning for target numerosities

When examining tuning stability across successive test periods, dlPFC neurons maintained stable numerosity tuning across all three test phases in which the matching numerosity could appear. Preferred numerosities and tuning widths remained largely unchanged over time compared to the sample numerosity (median Test1: $\sigma = 1.45$, median Test2: $\sigma = 1.84$, median Test3: $\sigma = 1.79$, Wilcoxon signed-rank test, Test1/Test2:

Z = -1.7842, p = 0.2975, Test1/Test3: Z = -1.9981, p = 0.1828, Test2/Test3: Z = -0.6416, p = 1, with Bonferroni correction, n = 68). In VIP, by contrast, tuning progressively deteriorated with each additional test stimulus. Population tuning curves became increasingly attenuated and asymmetric, and preferred numerosity representations shifted away from the initial sample numerosity.

Cross-correlation analyses of tuning curves confirmed these observations. In dlPFC, tuning similarity between sample and test phases remained high across all test periods (significantly higher than zero: $median_{test1}$ = 0.7091, $median_{test2}$ = 0.5518, $median_{test3}$ = 0.5285). In VIP, tuning similarity declined steadily and was no longer significantly different from zero by the third test phase ($median_{test1}$ = 0.5602, $median_{test2}$ = 0.2846, $median_{test3}$ = 0.3310). Consistent with this, population decoding accuracy in PFC remained stably above chance throughout all test periods (42 % ± 0.81, 43 % ± 0.70, and 40 % ± 0.73 during Test1, Test2, and Test3, respectively), whereas VIP decoding accuracy dropped to chance level in the final test phase (37 % ± 0.70, 37 % ± 0.71 and 32 % ± 0.64 during Test1, Test2, and Test3, respectively).

Together, these results demonstrate a clear functional dissociation between the two areas. VIP neurons primarily encode numerosity in a transient, stimulus-driven manner and fail to distinguish between relevant and irrelevant quantities over time. In contrast, dlPFC neurons selectively maintain and enhance representations of the sought-after numerosity, providing a stable neural correlate of recognition memory for numerical quantity.

3.2 Dynamic communication within the frontoparietal number network

From the same task and recordings as described above, the population-level interactions between the VIP and the dlPFC were investigated by using canonical correlation analysis (CCA) (Semedo et al., 2022). The analysis was restricted to multi-unit activity (MUA) during the fixation, sample, and delay periods to avoid contamination by motor-related signals. Only sessions with sufficient recording sites and balanced numbers of correct and error trials were included, resulting in 43 sessions used for population analyses.

3.2.1 Feedforward dominance after stimulus onset

During correct trials, population correlations between VIP and dlPFC were at chance level across all time lags during the fixation period. Shortly after sample onset, however,

canonical correlations increased sharply and became significantly greater than chance. Importantly, correlations were strongest at positive time lags, indicating that VIP activity leads subsequent dlPFC activity. This feedforward dominance emerged rapidly after stimulus onset and persisted for approximately 200 ms. During the later part of the sample period (400 – 700 ms), correlations remained significant but no longer showed a clear directional bias. Throughout the delay period, population correlations stayed above chance but were centered around zero lag, indicating balanced bidirectional communication without a consistent lead-lag relationship.

Centroid-based analyses of correlation peaks confirmed this temporal structure. Before neuronal responses emerged, correlation maxima were centered near zero lag. Once visual responses were present, maxima shifted toward positive lags of approximately +10 ms, consistent with feedforward signaling from VIP to dlPFC. At the transition from sample to delay, lag maxima returned to near zero and remained stable throughout the delay, reflecting symmetric inter-areal coupling.

Quantification of feedforward and feedback interactions further supported these observations. Both VIP-to-dlPFC feedforward correlations and dlPFC-to-VIP feedback correlations were significantly above chance from sample onset through the end of the delay. However, a significant feedforward dominance emerged specifically during the early sample period, when VIP-to-dlPFC correlations exceeded feedback correlations (until ~200 ms). This dominance vanished during later task phases, resulting in balanced bidirectional interaction. These interaction patterns were qualitatively similar across both monkeys, despite differences in absolute correlation strength.

3.2.2 Inter-areal communication is primarily driven by number neurons

To test whether inter-areal communication was specifically related to numerical processing, population correlations were analyzed separately for numerosity-selective (VIP: 151/257; dlPFC: 205/270) and non-selective neuronal populations (VIP: 106/257; dlPFC: 65/270). Recording sites were classified as sample-selective, delay-selective, or non-selective based on their numerosity tuning. VIP-dlPFC population correlations were strongest for sample-selective MUA and weaker, but still present, for delay-selective activity. In contrast, non-selective populations exhibited minimal correlations across time and lag. Mean population correlation values for sample-selective and delay-selective units significantly exceeded those of non-selective units ($p < 0.05$,

Mann–Whitney U test, two-sided), demonstrating that inter-areal communication was driven primarily by neurons encoding numerical information rather than by non-specific task-related activity.

3.2.3 Coordinated population activity breaks down during error trials

Error trials revealed a markedly different interaction pattern. As in correct trials, correlations during fixation were absent, and a transient increase in correlations with feedforward dominance occurred shortly after sample onset. However, during the later delay period, population correlations in error trials declined substantially and dropped below chance toward the end of the delay (below the 97.5^{th} percentile of the shuffled distribution). This breakdown occurred for both positive and negative lags, indicating a general loss of coordinated activity between VIP and dlPFC rather than a shift in interaction direction. Across both monkeys, correlation strength during correct trials consistently exceeded that during error trials, particularly at the end of the sample period and during the late delay.

Direct comparisons between correct and error trials showed that differences in population correlation strength were negligible during fixation but became significant during task-relevant epochs. Correlations were significantly stronger in correct trials across all lags toward the end of the sample (500 – 680 ms) and during the later delay phase (1260 – 1440 ms). Mean population correlation functions averaged across lags further confirmed that correct trials were characterized by sustained inter-areal coupling during these critical phases, whereas error trials showed weakened or disrupted interactions.

Together, these results demonstrate that accurate numerical performance is associated with dynamic and task-dependent communication between parietal and prefrontal neuronal populations. VIP exerts a transient feedforward influence on dlPFC during stimulus encoding, followed by sustained bidirectional interactions during working memory maintenance. This coordinated population activity is selectively driven by numerosity-encoding neurons and breaks down during error trials, highlighting the critical role of intact frontoparietal communication for successful abstract quantity judgments.

3.3 Comparative behavioral signatures of numerical cognition

In a further step, two crows were confronted with the same sequential delayed match-to-numerosity task described above. Both monkeys and crows successfully learned the task, achieving stable performance levels that were significantly above the 25 % chance level across all sessions (monkey 1: 27 sessions, 61.32 ± 1.36 %, $p < 0.001$; monkey 2: 33 sessions, 60.95 ± 0.58 %, $p < 0.001$; crow 1: 20 sessions, 59.90 ± 0.64 %, $p < 0.001$; crow 2: 20 sessions, 64.30 ± 1.14 %, $p < 0.001$; Binomial test). Behavioral performance functions revealed that, for all sample numerosities, both species reliably selected the matching numerosity more often than any nonmatching alternative. Accuracy was highest for numerosity 1 (peak: monkey 1, 95.02 ± 1.12 %; monkey 2, 96.01 ± 0.52 %; crow 1, 98.12 ± 0.47 %; crow 2, 98.02 ± 0.48 %) and declined progressively with increasing sample numerosity, reaching its lowest values for numerosity 4 (peak: monkey 1, 56.31 ± 1.96 %; monkey 2, 64.57 ± 1.99 %; crow 1, 66.08 ± 1.37 %; crow 2, 68.45 ± 1.63 %). In addition, both monkeys and crows exhibited clear numerical distance effects: nonmatching numerosities that were numerically adjacent to the target were most frequently confused with the sample, whereas more distant numerosities were rejected more reliably. Performance functions also broadened with increasing numerosity, reflecting a numerical size effect. Together, these behavioral signatures demonstrate that both species relied on approximate numerical representations consistent with the approximate number system (Merten and Nieder, 2009).

Task difficulty further increased with the number of intervening nonmatching test displays. Averaged across sample numerosities, performance declined as the number of nonmatches increased up to three in both monkeys and crows (three-way ANOVA with factors 'number of nonmatches', 'sample numerosity' and 'stimulus protocol'; monkey 1: $F(3,841) = 79.74$, $p < 0.001$; monkey 2: $F(3,1033) = 73.45$, $p < 0.001$; crow 1: $F(3,617) = 70.63$, $p < 0.001$; crow 2: $F(3,617) = 61.78$, $p < 0.001$). This effect was confirmed by significant main effects of trial type in all animals, indicating increasing demands on recognition memory and decision processes as more nonmatching numerosities had to be ignored. Performance within each trial type was also modulated by sample numerosity, as shown by significant interactions between trial type and numerosity. Numerosity 1 was consistently easiest for both species, whereas numerosity 4 posed the greatest challenge. Despite these effects, mean performance

across all trial types and numerosities remained significantly above chance in both species (for all sample numerosities and trial types, $p < 0.05$, Binomial test), demonstrating robust numerical recognition even under increased cognitive load.

3.3.1 Reaction-time differences between monkeys and crows

Despite these similar accuracy profiles, reaction-time analyses revealed striking differences between monkeys and crows. In monkeys, reaction times increased systematically with later test phases for all numerosities except, to some extent, numerosity 1. Responses were slowest when the matching numerosity appeared later in the test sequence, indicating cumulative processing costs across successive test displays (Pearson correlation, monkey 1, numerosity 1: $r = 0.17$, $p = 0.1357$; numerosity 2: $r = 0.79$, $p < 0.001$; numerosity 3: $r = 0.73$, $p < 0.001$; numerosity 4: $r = 0.79$, $p < 0.001$; monkey 2, numerosity 1: $r = 0.73$, $p < 0.001$; numerosity 2: $r = 0.81$, $p < 0.001$; numerosity 3: $r = 0.84$, $p < 0.001$; numerosity 4: $r = 0.82$, $p < 0.001$). Reaction times in monkeys were also strongly modulated by numerosity itself: responses were fastest for numerosity 1 and progressively slower for larger numerosities, with numerosity 4 eliciting the longest reaction times. These effects were highly consistent across sessions and animals, indicating that both test phase and numerical magnitude substantially influenced monkeys' decision latencies.

In contrast, crows showed the opposite temporal pattern. Reaction times decreased with later test phases, such that responses were fastest when the match appeared at later positions in the sequence. This speeding effect was observed for all numerosities in both crows, although it was weakest for numerosity 1 in one animal (Pearson correlation, crow 1, numerosity 1: $r = -0.78$, $p < 0.001$; numerosity 2: $r = -0.80$, $p < 0.001$; numerosity 3: $r = -0.86$, $p < 0.001$; numerosity 4: $r = -0.86$, $p < 0.001$; crow 2, numerosity 1: $r = -0.07$, $p = 0.6078$; numerosity 2: $r = -0.68$, $p < 0.001$; numerosity 3: $r = -0.63$, $p < 0.001$; numerosity 4: $r = -0.51$, $p < 0.001$). Moreover, unlike monkeys, crows' reaction times were largely invariant across numerosities: responses to numerosity 1 and numerosity 4 were of similar speed within each test phase. Thus, neither numerical magnitude nor increasing numbers of preceding nonmatches imposed measurable reaction-time costs in crows.

Linear regression analyses of reaction times across test phases confirmed these opposing patterns. Monkeys consistently exhibited positive slopes, indicating slower

responses with increasing test phase, whereas crows showed negative slopes, reflecting faster responses. These slopes differed highly significantly between species for all numerosities (Numerosity 1: $Z = -6.5638$, $p < 0.001$; Numerosity 2: $Z = -8.4107$, $p < 0.001$; Numerosity 3: Z -8.1540, $p < 0.001$; Numerosity 4: $Z = -8.3964$, $p < 0.001$, Wilcoxon rank sum test, two-sided). Within each species, slope steepness was qualitatively similar across numerosities, although effects were weakest for numerosity 1, especially in monkeys. These findings indicate fundamentally different temporal dynamics underlying numerical decisions in the two species.

To further quantify how strongly reaction times reflected task variables, a support vector machine classifier was trained to predict numerosity and test phase based solely on reaction times. Decoding accuracy was significantly above chance (8.33 %) in all animals, demonstrating that reaction times contained task-relevant information in both species. However, decoding accuracy was substantially higher in monkeys than in crows ($Z = -19.0613$, $p < 0.001$, Wilcoxon rank sum test, two-sided). In monkeys, confusion matrices showed clear structure along the main diagonal, indicating reliable classification of both numerosity and test phase. In crows, by contrast, decoding performance was weaker and lacked clear class-specific structure, suggesting that reaction times carried much less information about these task variables. Decoding accuracy did not differ between the two monkeys (monkey 1: 43.94 %, monkey 2: 44.84 %, $Z = -0.5394$, $p = 0.5896$, Wilcoxon rank sum test, two-sided) but differed between the two crows, reflecting individual variability within the species (crow 1: 32.68 %, crow 2: 27.10 %, $Z = 8.3495$, $p < 0.001$, Wilcoxon rank sum test, two-sided).

3.3.2 Accuracy in relation to reaction times

Finally, the relationship between reaction times and behavioral accuracy was examined. In monkeys, higher accuracy tended to be associated with faster responses, particularly in one animal, and no condition showed slower responses accompanying better performance. In crows, the opposite tendency was observed: higher accuracy was often associated with longer reaction times, especially in one animal. Importantly, faster responses in crows were not linked to reduced accuracy, arguing against a speed–accuracy tradeoff as an explanation for their rapid responses. Instead, these correlations indicate species-specific relationships between decision speed and performance.

Together, the results show that although monkeys and crows achieve comparable levels of numerical accuracy and exhibit similar psychophysical signatures of the approximate number system, their reaction-time dynamics differ noticeably. Monkeys' responses reflect increasing costs with larger numerosities and later test phases, whereas crows respond rapidly, with little influence of numerical magnitude and with speeding across successive test displays. These dissociations demonstrate that similar numerical performance can arise from fundamentally different cognitive strategies, as revealed by their distinct temporal decision profiles.

4. Discussion

Taken together, the three studies converge on a central insight: abstract numerical cognition emerges from distributed hierarchically organized neural circuits, whose dynamic coordination shape behavior. The results reveal specialization within the frontoparietal network, as well as species-specific strategies for processing numerical information. The following sections briefly discuss these findings to address their implications for neural coding, inter-areal communication and the evolutionary foundations of numerical cognition.

4.1 Behavioral numerosity representations and decision processes

In the here presented sequential match-to-numerosity task, both monkeys and crows showed decreasing accuracy with increasing numerosity, as well as greater confusion between numerically adjacent values, indicating clear evidence of both the numerical distance effect and the magnitude effect. These effects are classic hallmarks of the ANS (Nieder, 2018; Ditz & Nieder, 2016). The ANS contrasts with the OFS, which is believed to operate by tracking small numbers of up to four items by assigning individual markers to set items (Nieder, 2018; Ditz & Nieder, 2016). While the OFS provides precise number representations, it fails to do so for quantities greater than four. This system has been suggested for relative numerosity discrimination in human infants; however, monkeys show no signs of an OFS during absolute numerosity discrimination (Nieder, 2018). Several studies have reported the ANS in trained monkeys (Beran, 2007; Dehaene et al., 1998; Cantlon & Brannon, 2006), crows (Ditz & Nieder, 2016), in adult humans who are unable to count verbally (Dehaene et al., 1998) and in innumerate adults (Frank et al., 2008).

Crucially, the observed psychophysical properties remained consistent across different task variants, including sequential matching paradigms involving varying memory demands (up to three test numerosities were presented in this task). The persistence of the performance above chance, even under increased cognitive load, demonstrates that abstract numerical representations can be flexibly accessed and maintained across contexts, forming a stable basis for subsequent decision-making processes.

Although the accuracy profiles were similar across species, reaction-time analyses revealed that the decision dynamics were fundamentally different in monkeys and crows. Monkeys exhibited longer reaction times when faced with larger numerosities

and in later test phases, which suggests the presence of cumulative processing costs and possibly sequential comparison strategies. This resembles the way humans process information, whereby their performance slows when more items must be processed or when the target appears later in a sequence (Yakovlev et al., 2005).

By contrast, the reaction times of crows decreased across test phases, and their response speeds were largely numerosity-invariant, indicating a rapid, parallel and feedforward-like decision strategy. Their numerosity-to-action behavior, which was performed with very short response latencies, likewise suggests rapid stimulus–response mapping without significant memory involvement (Kirschhock & Nieder, 2022). These findings support the idea that crows rely on quick, low-cost perceptual strategies with less dependence on sequential checking or sustained working memory (Bobrowicz & Osvath, 2019).

Taken together, these divergences highlight that similar levels of numerical performance can emerge from fundamentally different decision-making mechanisms. This comparative pattern suggests that convergent cognitive abilities may rely on distinct neural architectures and computational strategies (Olkowicz et al., 2016; Nieder, 2021). Future research should examine whether these species differences persist under controlled temporal conditions, when distractors are present, or when working memory demands are increased.

4.2 Frontoparietal specialization for numerical recognition memory

At the neuronal level, the results demonstrate a clear functional dissociation between the parietal and prefrontal cortices (the primate core number network, Nieder & Miller, 2004) in processing numerical information, depending on whether or not the test stimulus matched the previously presented sample. dlPFC neurons exhibited only mild modification of neuronal activity when the same preferred numerosity was repeated. Specifically, there was no loss of selectivity in tuning when the same sample numerosity was presented again as a matching stimulus. This suggests that PFC neurons retain their numerosity selectivity between trial phases, even when interrupted or delayed by one or more nonmatching stimuli. However, the tuning selectivity of PFC neurons decreased specifically when the same numerosities were presented as nonmatches during the test periods. This means that, when numerosities are searched for as matches, PFC neurons respond with high selectivity and are finely tuned to them.

However, tuning and decoding accuracy significantly decrease when the same numerosities are not those memorized and attended to by the monkey. This distinction between target and irrelevant numerosities is a hallmark of recognition memory and directly supports decision-making processes, as it enables the selective evaluation of match versus nonmatch stimuli.

By contrast, the tuning selectivity of VIP neurons decreased significantly with repetitions of numerosities, regardless of whether they were presented as matching or nonmatching test stimuli. This suggests that VIP neurons do not distinguish between matching and nonmatching numerosities. Rather, they encode numerosity predominantly in a stimulus-driven and transient manner, with tuning emerging rapidly following stimulus presentation, but degrading across the delay and test phases. This is also consistent with the observed neuronal response latencies between VIP and dlPFC (see also Viswanathan & Nieder, 2013, 2015), as well as with established neuroanatomical pathways and functional connectivity patterns (Chafee & Goldman-Rakic, 2000; Crowe et al., 2013). This further reinforces the role of VIP in preliminary numerical integration prior to executive processing in the dlPFC.

Furthermore, differences in how the dlPFC and VIP neurons responded to repeated numerosities were observed. In our numerosity task, significant match enhancement effects, i.e. the increased neuronal activity that occurs when a stimulus matches a previously encountered stimulus stored in memory, were found only in the PFC. This active mechanism ensures that repetitive, irrelevant nonmatch stimuli are ignored (Engel & Wang, 2011). Match enhancement has previously been observed individually in the prefrontal (Miller et al., 1996; Freedman et al., 2003; Qi et al., 2012) but also parietal (Rawley & Constantinidis, 2010) cortices. Conversely, significant match suppression effects occurred in both the PFC and the VIP, with similar magnitudes in these two brain areas. This reduction may help PFC and VIP neurons to filter out familiar numerosities, enabling them to focus on novel stimuli. Together, match enhancement and match suppression create a balance that enables PFC neurons to efficiently process, store and respond to numerical quantity, enhancing the recognition of familiar numerosities while suppressing redundancy.

The observed neuronal patterns indicate that, although the PPC contributes to the initial encoding of numerical magnitude, the PFC is critical for maintaining, selecting, and prioritizing task-relevant information. This functional specialization supports a

hierarchical organization in which sensory-driven representations are transformed into goal-directed and decision-relevant signals.

Interestingly, these results contrast with those from a previous numerosity distractor task (Jacob & Nieder, 2014; Jacob et al., 2018). In the classic distractor task, an irrelevant stimulus is passively viewed for a set period of time. In our study, however, monkeys were presented with a variable number of sequential test displays, requiring them to actively evaluate each presentation to determine whether it matched the number of dots or not. The absence of distractors is a strength as it isolates the core processes involved, eliminating the additional complexity of attention-related factors. This difference in task design may explain why the dlPFC neuron tuning was affected by a passively viewed distractor (Jacob & Nieder, 2014; Jacob et al., 2016; Lin et al., 2023), whereas it remained stable in our study, in which unpredictable and sequential numerosity presentations required active evaluation and decision-making.

4.3 Inter-areal dynamics in the frontoparietal number network

In addition to local encoding specializations, numerical cognition depends on the dynamic interaction between the parietal and prefrontal regions. Population-level analyses using CCA revealed that inter-areal communication is strongly influenced by task demands and evolves over time. Building on previous research (Jacob et al., 2018), this new method moves beyond measuring simple synchronization strength to reveal the structure, temporal direction, and functional segregation of population-level communication between feedforward and feedback signals in frontoparietal circuits (see Semedo et al., 2022). Our results suggest that inter-areal population correlations are almost exclusively driven by numerosity-selective neurons, particularly sample-selective units, indicating that reported inter-areal communication is specific to numerical information rather than reflecting non-specific recurrent or bidirectional dynamics related to attention or task engagement.

During the initial stages of stimulus processing, the flow of information was predominantly influenced by feedforward signals from the VIP to the dlPFC, which is consistent with the transfer of sensory-derived numerical information. This transient dominance shortly after sample onset aligns with anatomical and physiological evidence suggesting that the IPS is the initial source of numerical information for the PFC (Nieder & Miller, 2004; Viswanathan & Nieder, 2013; Jacob & Nieder, 2014). As

the task progressed into delay periods, however, interactions became more symmetric, suggesting recurrent and bidirectional communication that supports working memory maintenance. This balance suggests that working memory relies on recurrent communication between the parietal and prefrontal cortex, a concept supported by models of distributed working memory. These models propose that prefrontal activity is stabilized and reinforced through reciprocal interactions with posterior cortical regions (Merchant et al., 2011; Christophel et al., 2017; Crowe et al., 2013; Genovesio et al., 2014).

Crucially, this coordinated coupling was disrupted during error trials, particularly in late delay phases, indicating that stable inter-areal communication is essential for accurate performance (a critical phase for transforming maintained representations into decision signals). Similar communication failures have been reported in frontoparietal networks during attentional lapses (Dotson et al., 2014) and impaired decision-making (Salazar et al., 2012). These findings highlight that numerical cognition emerges not only from numerosity-selective neurons per se but from temporally structured network dynamics that support the integration of sensory encoding, working memory, and decision-making processes within the frontoparietal system.

4.4 Conclusion

The findings of this thesis, based on three complementary studies, highlight how abstract numerical information is encoded, maintained and behaviorally deployed within and across species. At the neuronal level, representations of numerosity emerge in both the parietal and prefrontal cortex, but these regions contribute in functionally distinct ways. VIP neurons encode numerical magnitude in a stimulus-driven and transient manner, whereas dlPFC neurons selectively enhance representations of behaviorally relevant target numerosities. This supports recognition memory and directly contributes to goal-directed decision-making by linking numerical representations to behavioral choices. Population-level analyses further reveal that successful numerical judgements depend on dynamic, task-specific communication between the parietal and prefrontal areas. Specifically, early feedforward influences from the parietal cortex to the prefrontal cortex during stimulus encoding transition to sustained bidirectional interactions during working memory maintenance; this coordinated coupling is disrupted during error trials.

At the behavioral level, both primates and corvids exhibit hallmark signatures of the approximate number system, including numerical distance and size effects. This demonstrates the convergent evolution of abstract quantity processing. However, differences in reaction time dynamics reveal striking species-specific variations, suggesting that comparable numerical performance can arise from distinct cognitive strategies and underlying decision-making mechanisms.

5. References

Agrillo, C. & Bisazza, A. (2018). Understanding the origin of number sense: a review of fish studies. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1740).

Anderson, B., Mruczek, R. E., Kawasaki, K. & Sheinberg, D. (2008). Effects of familiarity on neural activity in monkey inferior temporal lobe. *Cerebral cortex (New York*, *N.Y. : 1991)*, *18*(11), 2540–2552.

Anderson, M. C. & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*(6826):366.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044):556–559.

Barbas, H. & Mesulam, M. M. (1985). Cortical afferent input to the principals region of the rhesus monkey. *Neuroscience*, *15*(3), 619-637.

Bobrowicz, K., & Osvath, M. (2019). Cognition in the fast lane: Ravens' gazes are half as short as humans' when choosing objects. *Animal Behavior and Cognition*, *6*(2), 81–97.

Brown, M. W. & Aggleton, J. P. (2001). Recognition memory: what are the roles of the perirhinal cortex and hippocampus?. *Nature Reviews Neuroscience*, *2*(1): 51-61.

Buschman, T. J., & Miller, E. K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science*, *315*(5820), 1860-1862.

Cantlon, J. F. & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological science*, *17*(5): 401–406.

Castner, S. A., Goldman-Rakic, P. S., & Williams, G. V. (2004). Animal models of working memory: insights for targeting cognitive dysfunction in schizophrenia. *Psychopharmacology*, *174*: 111–125.

Cavada, C. & Goldman-Rakic, P. S. (1989). Posterior parietal cortex in rhesus monkey: Ii. evidence for segregated corticocortical networks linking sensory and limbic areas with the frontal lobe. *Journal of Comparative Neurology*, *287*(4): 422–445.

Chafee, M. V. & Goldman-Rakic, P. S. (2000). Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *Journal of neurophysiology*, *83*(3): 1550–1566.

Chafee, M. V. & Goldman-Rakic, P. S. (1998). Matching patterns of activity in primate prefrontal area 8a and parietal area 7ip neurons during a spatial working memorytask. *Journal of neurophysiology*, *79*(6): 2919-2940.

Chan, L. K. & Hayward, W. G. (2013). Visual search. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(4), 415-429.

Chelazzi, L., Miller, E. K., Duncan, J., and Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427):345–347.

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The Distributed Nature of Working Memory. *Trends in cognitive sciences*, *21*(2), 111–124.

Constantinidis, C. & Steinmetz, M. A. (1996). Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *Journal of Neurophysiology*, *76*(2):1352–1355.

Crowe, D. A., Goodwin, S. J., Blackman, R. K., Sakellaridi, S., Sponheim, S. R., MacDonald III, A. W., & Chafee, M. V. (2013). Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nature Neuroscience*, *16*(10):1484–1491.

Dash, P. K., Moore, A. N., Kobori, N., & Runyan, J. D. (2007). Molecular activity underlying working memory. *Learning and memory*, *14*(8): 554–563.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1-2):1–42.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, L. (1998). Abstract representations of numbers in the animal and human brain. *Trends in neurosciences*, *21*(8):355–361.

Dotson, N. M., Salazar, R. F., & Gray, C. M. (2014). Frontoparietal correlation dynamics reveal interplay between integration and segregation during visual working memory. *Journal of neuroscience*, *34*(41), 13600–13613.

di Pellegrino, G. & Wise, S. P. (1993). Visuospatial versus visuomotor activity in the premotor and prefrontal cortex of a primate. *Journal of Neuroscience*, *13*(3): 1227–1243.

Ditz, H. M., & Nieder, A. (2016). Numerosity representations in crows obey the Weber–Fechner law. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1827).

Dorris, M. C. & Glimcher, P. W. (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron*, *44*(2): 365–378.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive psychology*, *30*(3): 257–303.

Engel, T. A. & Wang, X. J. (2011) Same or different? A neural circuit mechanism of similarity-based pattern match decision making. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *31*(19): 6982–6996.

Everling, S., Tinsley, C. J., Gaffan, D., & Duncan, J. (2002). Filtering of neural signals by focused attention in the monkey prefrontal cortex. *Nature neuroscience*, *5*(7), 671-676.

Fanini, A. & Assad, J. A. (2009). Direction selectivity of neurons in the macaque lateral intraparietal area. *Journal of neurophysiology*, *101*(1): 289–305.

Fechner, G. (1860). *Elemente der psychophysik* (vol. iii). Breitkopf and Härtel.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in cognitive sciences*, *8*(7): 307–314.

Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., & Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proceedings of the National Academy of Sciences*, *108*(42): 17510-17515.

Fitzgerald, J. K., Freedman, D. J., & Assad, J. A. (2011). Generalized associative representations in parietal cortex. Nature neuroscience, 14(8): 1075–1079.

Frank, M. C., Everett, D. L., Fedorenko, E., & Gibson, E. (2008). Number as a cognitive technology: Evidence from pirahã language and cognition. *Cognition*, *108*(3): 819–824.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *Journal of Neuroscience*, *23*(12), 5235-5246.

Freedman, D. J. & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature*, *443*(7107): 85–88.

Freedman, D. J. & Assad, J. A. (2011). A proposed common neural mechanism for categorization and perceptual decisions. *Nature neuroscience*, *14*(2): 143–146.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2006). Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. *Cerebral cortex (New York*, *N.Y. : 1991)*, *16*(11): 1631–1644.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, *291*(5502): 312– 316.

Fuster, J. M. (2008). *The prefrontal cortex* (4th ed.). Academic Press.

Gazzola, A., Vallortigara, G., & Pellitteri-Rosa, D. (2018). Continuous and discrete quantity discrimination in tortoises. *Biology Letters*, *14*(12).

Genovesio, A., Wise, S. P., & Passingham, R. E. (2014). Prefrontal-parietal function: from foraging to foresight. *Trends in cognitive sciences*, *18*(2), 72–81.

Gold, J. I. & Shadlen, M. N. (2000). Representation of a perceptual decision in developing oculomotor commands. *Nature*, *404*(6776): 390–394.

Gold, J. I. & Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, *30*(1): 535–574.

Goldman-Rakic, P. S. (1988). Topography of cognition: parallel distributed networks in primate association cortex. *Annual review of neuroscience*, *11*(1): 137–156.

Goldman-Rakic, P. S. (1991). Cellular and circuit basis of working memory in prefrontal cortex of nonhuman primates. *Progress in brain research*, *85*, 325–336.

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, *14*(3):477–485.

Grant, D. S. (1976). Effect of sample presentation time on long-delay matching in the pigeon. *Learning and Motivation*, *7*(4): 580–590.

Grefkes, C. & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of anatomy*, *207*(1): 3–17.

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, *520*(7546): 220–223.

Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Greentree, A. D., & Dyer, A. G. (2018). Numerical ordering of zero in honey bees. *Science*, *360*(6393): 1124–1126.

Jacob, S. N. & Nieder, A. (2014). Complementary roles for primate frontal and parietal cortex in guarding working memory from distractor stimuli. *Neuron*, *83*(1): 226-237.

Jacob, S. N., Hähnke, D., & Nieder, A. (2018). Structuring of Abstract Working Memory Content by Fronto-parietal Synchrony in Primate Cortex. *Neuron*, *99*(3), 588–597.e5.

Jacob, S. N., Stalter, M., & Nieder, A. (2016). Cell-type-specific modulation of targets and distractors by dopamine D1 receptors in primate prefrontal cortex. *Nature communications*, *7*(1), 13218.

Jarvik, M., Goldfarb, T., & Carley, J. (1969). Influence of interference on delayed matching in monkeys. *Journal of Experimental Psychology*, *81*(1): 1.

Katsuki, F. & Constantinidis, C. (2012a). Early involvement of prefrontal cortex in visual bottom-up attention. *Nature neuroscience*, *15*(8): 1160.

Katsuki, F., & Constantinidis, C. (2012b). Unique and shared roles of the posterior parietal and dorsolateral prefrontal cortex in cognitive functions. *Frontiers in integrative neuroscience*, *6*, 17.

Kirschhock, M. E., & Nieder, A. (2022). Number selective sensorimotor neurons in the crow translate perceived numerosity into number of actions. *Nature communications*, *13*(1), 6913.

Krawczyk, D. C. (2002). Contributions of the prefrontal cortex to the neural basis of human decision making. *Neuroscience and Biobehavioral Reviews*, *26*(6): 631–664.

Kusunoki, M., Sigala, N., Gaffan, D., & Duncan, J. (2009). Detection of fixed and variable targets in the monkey prefrontal cortex. *Cerebral Cortex*, *19*(11): 2522–2534.

Lara, A. H. & Wallis, J. D. (2015). The role of prefrontal cortex in working memory: a mini review. *Frontiers in systems neuroscience*, *9*:173.

Lennert, T. & Martinez-Trujillo, J. (2011). Strength of response suppression to distracter stimuli determines attentional-filtering performance in primate prefrontal neurons. *Neuron*, *70*(1): 141–152.

Lennert, T., Cipriani, R., Jolicoeur, P., Cheyne, D., & Martinez-Trujillo, J. C. (2011). Attentional modulation of neuromagnetic evoked responses in early human visual cortex and parietal lobe following a rank-order rule. *Journal of Neuroscience*, *31*(48): 17622– 17636.

Leon, M. I. & Shadlen, M. N. (1998). Exploring the neurophysiology of decisions. *Neuron*, *21*(4):669–672.

Lim, S., McKee, J. L., Woloszyn, L., Amit, Y., Freedman, D. J., Sheinberg, D. L., & Brunel, N. (2015). Inferring learning rules from distributions of firing rates in cortical neurons. *Nature neuroscience*, *18*(12): 1804–1810.

Lin, F.-C., Whiting, M. J., Hsieh, M.-Y., Shaner, P.-J. L., & Lin, S.-M. (2021). Superior continuous quantity discrimination in a freshwater turtle. *Frontiers in Zoology*, *18*: 1–11.

Lin, X. X., Nieder, A., & Jacob, S. N. (2023). The neuronal implementation of representational geometry in primate prefrontal cortex. *Science advances*, *9*(50), eadh8685.

Maki, W. S., Moe, J. C., & Bierley, C. M. (1977). Short-term memory for stimuli, responses, and reinforcers. Journal of Experimental Psychology: *Animal Behavior Processes*, *3*(2): 156.

Mandler, G. & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of experimental psychology: general*, *111*(1): 1.

Manes, F., Sahakian, B., Clark, L., Rogers, R., Antoun, N., Aitken, M., & Robbins, T. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain*, *125*(3): 624–639.

Meister, M. L., Hennig, J. A., & Huk, A. C. (2013). Signal multiplexing and single neuron computations in lateral intraparietal area during decision-making. *Journal of Neuroscience*, *33*(6): 2254–2267.

Merchant, H., Crowe, D. A., Robertson, M. S., Fortes, A. F., & Georgopoulos, A. P. (2011). Top-down spatial categorization signal from prefrontal to posterior parietal cortex in the primate. *Frontiers in systems neuroscience*, *5*, 69.

Merten, K. & Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. *Journal of Cognitive Neuroscience*, *21*(2): 333–346.

Merten, K. & Nieder, A. (2012). Active encoding of decisions about stimulus absence in primate prefrontal cortex neurons. *Proceedings of the National Academy of Sciences*, *109*(16): 6289–6294.

Merten, K. & Nieder, A. (2013). Comparison of abstract decision encoding in the monkey prefrontal cortex, the presupplementary, and cingulate motor areas. *Journal of Neurophysiology*, *110*(1): 19–32.

Miles, R. C. (1971). Species differences in “transmitting” spatial location information. In *Cognitive processes of nonhuman primates* (pp. 165-179). Academic Press.

Miller, E. K. & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual review of neuroscience*, *24*(1): 167-202.

Miller, E. K. & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, *263*(5146): 520-522.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of neuroscience*, *16*(16): 5154–5167.

Milner, B. (1963). Effects of different brain lesions on card sorting: The role of the frontal lobes. *Archives of neurology*, *9*(1): 90–100.

Mishkin, M. & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research*, *6*(1): 57– 77.

Mishkin, M., Vest, B., Waxler, M., & Rosvold, H. E. (1969). A re-examination of the effects of frontal lesions on object alternation. *Neuropsychologia*, *7*(4): 357–363.

Moyer, R. S. & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109): 1519-1520.

Murray, J. D., Jaramillo, J., & Wang, X.-J. (2017). Working memory and decision-making in a frontoparietal circuit model. *Journal of Neuroscience*, *37*(50): 12167–12186.

Müller, H. J. & Krummenacher, J. (2006). Visual search and selective attention. *Visual Cognition*, *14*(4-8): 389–410.

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences*, *10*9(29): 11860–11865.

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6): 366-382.

Nieder, A. (2018). Evolution of cognitive & neural solutions enabling numerosity judgements: lessons from primates and corvids. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1740).

Nieder, A. (2020). The Adaptive Value of Numerical Competence. *Trends in ecology and evolution*, *35*(7): 605–617.

Nieder A. (2021). The Evolutionary History of Brains for Numbers. *Trends in cognitive sciences*, *25*(7), 608–621.

Nieder, A. & Merten, K. (2007). A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *Journal of Neuroscience*, *27*(22): 5986–5993.

Nieder, A. & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, *37*(1): 149–157.

Nieder, A. & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences*, *101*(19): 7457–7462.

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, *297*(5587): 1708–1711.

Okuyama, S., Kuki, T., & Mushiake, H. (2015). Representation of the numerosity 'zero'in the parietal cortex of the monkey. *Scientific Reports*, *5*(1): 10059.

Olkowicz, S., Kocourek, M., Lučan, R. K., Porteš, M., Fitch, W. T., Herculano-Houzel, S., & Němec, P. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(26), 7255–7260.

Pasternak, T. & Zaksas, D. (2003). Stimulus specificity and temporal dynamics of working memory for visual motion. *Journal of Neurophysiology*, *90*(4): 2757–2762.

Platt, M. L. & Glimcher, P. W. (1999). Neural correlates of decision variables in parietal cortex. *Nature*, *400*(6741): 233–238.

Qi, X., Katsuki, F., Meyer, T., Rawley, J. B., Zhou, X., Douglas, K. L., & Constantinidis, C. (2010). Comparison of neural activity related to working memory in primate dorsolateral prefrontal and posterior parietal cortex. *Frontiers in systems neuroscience*, *4*:12.

Qi, X.-L., Meyer, T., Stanford, T. R., & Constantinidis, C. (2012). Neural correlates of a decision variable before learning to perform a match/non-match task. *Journal of Neuroscience*, *32*(18): 6161–6169.

Quintana, J. & Fuster, J. M. (1992). Mnemonic and predictive functions of cortical neurons in a memory task. *Neuroreport*, *3*(8): 721–724.

Quintana, J., Fuster, J. M., & Yajeya, J. (1989). Effects of cooling parietal cortex on prefrontal units in delay tasks. *Brain research*, *503*(1): 100–110.

Rainer, G. & Miller, E. K. (2000). Effects of visual experience on the representation of objects in the prefrontal cortex. *Neuron*, *27*(1): 179-189.

Ramirez-Cardenas, A., Moskaleva, M., & Nieder, A. (2016). Neuronal representation of numerosity zero in the primate parieto-frontal number network. *Current Biology*, *26*(10): 1285–1294.

Rawley, J. B. & Constantinidis, C. (2009). Neural correlates of learning and working memory in the primate posterior parietal cortex. *Neurobiology of learning and memory*, *91*(2): 129–138.

Rawley, J. B. & Constantinidis, C. (2010). Effects of task and coordinate frame of attention in area 7a of the primate posterior parietal cortex. *Journal of vision*, *10*(1): 12– 12.

Riches, I. P., Wilson, F. A., & Brown, M. W. (1991). The effects of visual stimulation and memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. *Journal of Neuroscience*, *11*(6): 1763-1779.

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS biology*, *5*(8): e208.

Romo, R., & Salinas, E. (2001). Touch and go: decision-making mechanisms in somatosensation. *Annual review of neuroscience*, *24*(1): 107-137.

Rudebeck, P. H., Saunders, R. C., Lundgren, D. A., & Murray, E. A. (2017). Specialized representations of value in the orbital and ventrolateral prefrontal cortex: desirability versus availability of outcomes. *Neuron*, *95*(5):1208–1220.

Rugani, R., Regolin, L., & Vallortigara, G. (2008). Discrimination of small numerosities in young chicks. Journal of experimental psychology: *Animal behavior processes*, *34*(3): 388.

Rushworth, M. F., Nixon, P. D., Eacott, M. J., & Passingham, R. E. (1997). Ventral prefrontal cortex is not essential for working memory. *Journal of Neuroscience*, *17*(12): 4829–4838.

Sakai, K., Rowe, J. B., & Passingham, R. E. (2002). Active maintenance in prefrontal area 46 creates distractor-resistant memory. *Nature neuroscience*, *5*(5): 479–484.

Salazar, R., Dotson, N., Bressler, S., & Gray, C. (2012). Content-specific fronto-parietal synchronization during visual working memory. *Science*, *338*(6110): 1097–1100.

Sawamura, H., Shima, K., & Tanji, J. (2002). Numerical representation for action in the parietal cortex of the monkey. *Nature*, *415*(6874): 918.

Schall, J. D. (2001). Neural basis of deciding, choosing and acting. Nature Reviews *Neuroscience*, *2*(1): 33–42.

Semedo, J. D., Jasper, A. I., Zandvakili, A., Krishna, A., Aschner, A., Machens, C. K., Kohn, A., & Yu, B. M. (2022). Feedforward and feedback interactions between visual cortical areas use different population activity patterns. *Nature communications*, *13*(1): 1099.

Shadlen, M. N. & Newsome, W. T. (1996). Motion perception: seeing and deciding. *Proceedings of the national academy of sciences*, *93*(2): 628–633.

Shettleworth, S. J. (2010). Cognition, evolution, and behavior. *Oxford University Press*.

Siegel, M., Buschman, T. J., & Miller, E. K. (2015). Cortical information flow during flexible sensorimotor decisions. *Science*, *348*(6241): 1352–1355.

Siegler, R. S. & Booth, J. L. (2004). Development of numerical estimation in young children. *Child development*, *75*(2): 428–444.

Stancher, G., Rugani, R., Regolin, L., & Vallortigara, G. (2015). Numerical discrimination by frogs (bombina orientalis). *Animal Cognition*, *18*: 219–229.

Sugrue, L. P., Corrado, G. S., & Newsome, W. T. (2004). Matching behavior and the representation of value in the parietal cortex. *Science*, *304*(5678): 1782–1787.

Suzuki, M. & Gottlieb, J. (2013). Distinct neural mechanisms of distractor suppression in the frontal and parietal lobe. *Nature neuroscience*, *16*(1): 98.

Tamura, K., Takeda, M., Setsuie, R., Tsubota, T., Hirabayashi, T., Miyamoto, K., & Miyashita, Y. (2017). Conversion of object identity to object-general semantic value in the primate temporal cortex. *Science*, *357*(6352): 687–692.

Tanji, J. & Hoshi, E. (2008). Role of the lateral prefrontal cortex in executive behavioral control. *Physiological reviews*, *88*(1): 37–57.

Toth, L. J. & Assad, J. A. (2002). Dynamic coding of behaviourally relevant stimuli in parietal cortex. *Nature*, *415*(6868): 165–168.

Treisman, A. M. & Gelade, G. (1980). A feature-integration theory of attention. Cognitive psychology, 12(1): 97–136.

Tudusciuc, O. & Nieder, A. (2009). Contributions of primate prefrontal and posterior parietal cortices to length and numerosity representation. *Journal of Neurophysiology*, *101*(6): 2984–2994.

Viswanathan, P. & Nieder, A. (2013). Neuronal correlates of a visual “sense of number” in primate parietal and prefrontal cortices. *Proceedings of the National Academy of Sciences*, *110*(27): 11187–11192.

Viswanathan, P., & Nieder, A. (2015). Differential impact of behavioral relevance on quantity coding in primate frontal and parietal neurons. *Current biology : CB*, *25*(10), 1259–1269.

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*(7067): 500.

Wang, X.-J. (2008). Decision making in recurrent neuronal circuits. Neuron, 60(2):215–234. Whitlock, J. R. (2017). Posterior parietal cortex. *Current biology*, *27*(14): R691–R695.

Whitlock, J. R. (2017). Posterior parietal cortex. *Current biology*, *27*(14): R691-R695.

Winters, B. D., Saksida, L. M., & Bussey, T. J. (2008). Object recognition memory: neurobiological mechanisms of encoding, consolidation and retrieval. *Neuroscience and biobehavioral reviews*, *32*(5): 1055–1070.

Woloszyn, L. & Sheinberg, D. L. (2012). Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. *Neuron*, *74*(1): 193–205.

Xiang, J. Z. & Brown, M. W. (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. *Neuropharmacology*, *37*(4-5): 657–676.

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental science*, *8*(1): 88–101.

Yakovlev, V., Bernacchia, A., Orlov, T., Hochstein, S., & Amit, D. (2005). Multi-item working memory--a behavioral study. *Cerebral cortex (New York*, *N.Y. : 1991)*, *15*(5), 602–615.

Yang, T. & Shadlen, M. N. (2007). Probabilistic reasoning by neurons. *Nature*, *447*(7148): 1075–1080.

Zaksas, D. & Pasternak, T. (2006). Directional signals in the prefrontal cortex and in area mt during a working memory for visual motion task. *Journal of Neuroscience*, *26*(45): 11726–11742.

Zhou, Y., Rosen, M. C., Swaminathan, S. K., Masse, N. Y., Zhu, O., & Freedman, D. J. (2021). Distributed functions of prefrontal and parietal cortices during sequential categorical decisions. *Elife*, *10*: e58782.

Part II.

Individual Studies

Publication 1: Neuronal encoding of recognition memory for numerical quantities in macaque intraparietal and prefrontal cortices.

This chapter is identical to the published version of:

Machts, T., Grüb, J., & Nieder, A. (2025). Neuronal encoding of recognition memory for numerical quantities in macaque intraparietal and prefrontal cortices. *Progress in Neurobiology*, *251*, 102794. https://doi.org/10.1016/j.pneurobio.2025.102794

Progress in Neurobiology 251 (2025) 102794

Contents lists available at ScienceDirect

Progress in Neurobiology

journal homepage: www.elsevier.com/locate/pneurobio

Neuronal encoding of recognition memory for numerical quantities in macaque intraparietal and prefrontal cortices

Tobias Machts, Julia Grüb, Andreas Nieder [1,*]

Animal Physiology, Institute of Neurobiology, University of Tuebingen, Auf der Morgenstelle 28, Tuebingen 72076, Germany

ARTICLE INFO

Keywords:
Recognition memory
Single neurons
Monkey
Prefrontal cortex
Intraparietal cortex

ABSTRACT

The parieto-frontal number network in primates is vital for extracting and memorizing numerical information. However, how neurons in these regions retain abstract numerical categories to recognize target numbers amidst ongoing numerical input is unclear. To explore this, single neurons were recorded from the ventral intraparietal cortex (VIP) and lateral prefrontal cortex (PFC) of two male macaques trained to memorize and recognize target numerosities while viewing sequences of irrelevant numerosities. In the VIP, neuronal selectivity for both target and irrelevant numerosities declined rapidly, making it unable to distinguish relevant from irrelevant quantities. Conversely, PFC neurons maintained selective tuning for target numerosities over time but not for irrelevant ones, enabling the distinction between sought and irrelevant quantities. Match enhancement effects, where firing increased for repeated target numerosities, were observed only in the PFC. In contrast, match suppression effects, involving reduced firing for repeated target numerosities, occurred in both the VIP and PFC. These findings suggest the VIP primarily encodes displayed numerosities, while the PFC is specialized for processing, storing, and recognizing numerical quantities by enhancing familiar numerosities. This highlights the PFC's key role in recognition memory, contrasting with the transient coding observed in the VIP.

1. Introduction

We usually encounter important objects or events not only once in our lives, but several times. The cognitive function that enables us to recognize that something currently perceived has been experienced before is called recognition memory, a critical component of explicit memory (Winters et al., 2008). While recognition memory has been extensively studied for arbitrary sensory stimuli (Brown and Aggleton, 2001), its exploration in the context of abstract categories, such as numbers, has been neglected. Recognition memory of numbers is essential for tasks such as remembering item prices or accurately dosing medications. Recognition memory of numerosities — the number of items in a set — also enables animals to efficiently navigate their environment, make informed decisions, and engage in behaviors critical for survival and reproduction (Nieder, 2020).

Studies of sensory recognition memory report neurons whose activity indicates whether a stimulus is familiar or novel (Rainer and Miller, 2000; Brown and Aggleton, 2001). Most studies have identified neurons in high-level cortical areas whose response magnitudes decrease with repetition and thus familiarity, a phenomenon known as 'repetition suppression' (Fahy et al., 1993; Li et al., 1993; Miller and Desimone, 1994; Riches et al., 1991; Xiang and Brown, 1998; Ng et al., 2014). However, other mechanisms for signaling familiarity have also been observed (Meyer and Rust, 2018), such as more selective tuning (Rainer and Miller, 2000; Freedman et al., 2006; Anderson et al., 2008;), tuning peak enhancement (Woloszyn and Sheinberg, 2012; Lim et al., 2015), or even increases in firing rate (Tamura et al., 2017). Understanding which signals correlate with perceptual memories and how these neuronal correlates differ along the cortical processing hierarchy remains an area of active research.

While sensory recognition memory has been studied using arbitrary pictogram stimuli, the memory necessary for recognizing abstract categories, such as numerical quantities, in a sequence of numerosities has not been investigated. The special feature of numerical quantities in contrast to arbitrary visual stimuli is that they are semantically related to each other by ordinal magnitude (e.g., three is greater than two but less than four). This relationship is often described using the metaphor of the 'mental number line,' where numbers are arranged in an ordered

* Corresponding author.
E-mail address: andreas.nieder@uni-tuebingen.de (A. Nieder).
[1] ORCID: 0000-0001-6381-0375

https://doi.org/10.1016/j.pneurobio.2025.102794
Received 6 December 2024; Received in revised form 1 May 2025; Accepted 10 June 2025
Available online 13 June 2025

sequence (Galton, 1880; Dehaene et al., 1993; Zorzi et al., 2002; Nieder, 2020). Recognition memory of numerical quantities therefore heavily depends on the representation of, or tuning to, target numerosities as opposed to neighboring non-target numerosities as a unique feature of parameterized quantities.

A key behavioral phenomenon related to numerical cognition based on the number line is the 'numerical distance effect' (Moyer and Landauer, 1967; Buckley and Gillman, 1974). This effect highlights that it is more challenging to differentiate between numbers that are close together (e.g., 2 and 3) compared to those that are further apart (e.g., 2 and 8). This effect is mirrored in the neurons' response properties. In the parieto-frontal number network of primates, numerosity-selective neurons are specifically responsive to the presentation and short-term memorization of particular numerical values (Nieder et al., 2002, 2006; Sawamura et al., 2002; Nieder and Miller, 2003; 2004; Nieder and Merten, 2007; Okuyama et al., 2015; Nieder, 2016). Like sensory and motor neurons that are tuned to a specific feature or parameterized coding space, numerosity-selective neurons exhibit a bell-shaped tuning curve. This means they have a maximum firing rate for their preferred numerosity and show progressively lower activity as the presented numbers deviate from this preferred value. The amplitude and width of these tuning curves reflect the neurons' selectivity for specific numerosities (Nieder, 2023). This tuning selectivity is a key aspect not only of sensory numerosity representations but also for the maintenance of numerical recognition memory; neurons whose tuning deteriorates with time and numerosity repetition would be unsuitable as a correlate of recognition memory for numerosity.

In this study, we investigated where and how numerosity-selective neurons in the parieto-frontal number network can serve as a neuronal representation of recognition memory for numerical quantity. Both the areas within the intraparietal sulcus (IPS) and the lateral prefrontal cortex (PFC) have been repeatedly found to encode numerical information, forming the core number network in primates (Nieder, 2024). In contrast, such information is largely absent in the anterior inferotemporal cortex (Nieder and Miller, 2004).

We trained two rhesus macaques to perform a sequential delayed-match-to-numerosity task with variable numbers of non-matching visual numerosity stimuli between sample and matching test displays. The delayed match-to-numerosity task involves a working memory component that interacts closely with recognition memory (Miller et al., 1996). While working memory temporarily stores and manipulates information, recognition memory retrieves and identifies previously encountered information. Both memory systems are crucial for accurate decision-making and memory-based tasks, often working together to support effective cognitive processing.

The monkeys were required to form an internal representation or template, akin to a search image, for recognizing and identifying specific cued numerical quantities. While the monkeys performed this task, we simultaneously recorded single-neuron activity from the lateral PFC and the ventral intraparietal area (VIP) in the intraparietal cortex, two key number processing areas in the primate brain (Nieder and Miller, 2004; Viswanathan and Nieder, 2013; 2015,; Nieder and Dehaene, 2009; Nieder, 2016). To explore the neuronal correlate for recognition memory of numerical quantity categories, we quantified and compared the numerosity tuning curve characteristics and population decoding accuracies for repeated presentations of both matching and non-matching numerosities in the two cortical areas under the exact same recording conditions.

2. Results

2.1. Behavioral performance

Two monkeys were trained on a modified version of the delayed match-to-numerosity task. They were presented with sample numerosities 1–4, followed by a delay period, and then exposed in equal proportions of up to four different sequential test phases (see Materials and Methods; Fig. 1A). The monkeys were required to memorize the variable sample numerosity and to search the sequence of test displays for a numerosity that matched the sample numerosity, and they were rewarded for responding whenever one of the displays in the test periods showed a matching numerosity (correct trials). This encouraged them to inspect each test stimulus and to keep searching for the matching numerosity.

Both monkeys mastered the task, with mean performances across sessions significantly exceeding the chance level of 25 % (monkey 1: 27 sessions, 61.32 ± 1.36 %, $p < 0.001$; monkey 2: 33 sessions, 60.95 ± 0.58 %, $p < 0.001$, Binomial test) (see Fig. S1 for chance level calculation). To evaluate the monkeys' ability to identify the correct matching numerosity while disregarding the nonmatching numerosities, behavioral performance functions were computed for each session. These functions reflect the probability that a monkey judged the numerosities in the test periods as equal to the sample numerosity. For all sample numerosities, both monkeys more frequently selected the matching numerosity over the nonmatching numerosities (Fig. 1B,D). The best performance was achieved for numerosity 1 (peak: monkey 1, 95.02 ± 1.12 %; monkey 2, 96.01 ± 0.52 %) and declined with increasing numerosity up to numerosity 4 (peak: monkey 1, 56.31 ± 1.96 %; monkey 2, 64.57 ± 1.99 %). As expected, based on the numerical distance effect, which suggests that neighboring numerosities are the most challenging to discriminate (Nieder and Miller, 2003; Merten and Nieder, 2009), both monkeys exhibited the highest error rates for nonmatching numerosities adjacent to the respective target numerosity, with noticeable improvements for more distant nonmatch numerosities. Additionally, the performance functions systematically widened (i.e., became less selective) for increasing target numerosities, reflecting the numerical size effect, which indicates that, at a given numerical distance, discriminating target numerosities becomes more difficult as their values increase (Nieder, 2020). Thus, the monkeys' performance exhibits the characteristics of the approximate number system (ANS), a non-symbolic estimation system that also emerges in humans when they are prevented from counting and rely on non-symbolic numerosity estimation (Merten and Nieder, 2009; Nieder, 2024).

We computed the averaged performance for each trial type (i.e., no nonmatch, one nonmatch, two nonmatches, three nonmatches) across the four sample numerosities (Fig. 1C,E). As anticipated, the number of inserted nonmatch numerosities influenced the monkeys' ability to recall and detect the matching numerosity (two-way ANOVA with the factors 'number of nonmatches' and 'sample numerosity'; monkey 1: $F(3431) = 60.1$, $p < 0.001$; monkey 2: $F(3527) = 66.49$, $p < 0.001$). Compared to the pure matching trials in which no nonmatch numerosity was shown, the monkeys' overall performance progressively decreased with the number of nonmatches shown in a trial. The observed decline in behavioral performance with increasing nonmatch trials reflects the growing cognitive demands on recognition memory, a critical aspect of the study. Monkey 1, but not Monkey 2, shows a mild, idiosyncratic performance increase for numerosities following three nonmatches. However, this pattern is uncorrelated with Monkey 2's behavior and likely reflects an individual, unknown strategy. Furthermore, behavioral performance depended on the presented sample numerosity within each trial type (significant interaction of trial type and presented sample numerosity; monkey 1 $F(9, 431) = 5.16$, $p < 0.001$; monkey 2 $F(9, 527) = 4.27$, $p < 0.001$). As a reflection of the numerical size effect, (balanced) trials in which the monkeys had to search for numerosity 1 were the easiest, while trials in which numerosity 4 was the target challenged the monkeys the most (Fig. 1**C,E**). Nevertheless, the averaged performances for all trial types and sample numerosities were significantly higher than chance (for all sample numerosities and trial types, $p < 0.05$, Binomial test), revealing that the monkeys were able to find the correct numerosity most of the time.

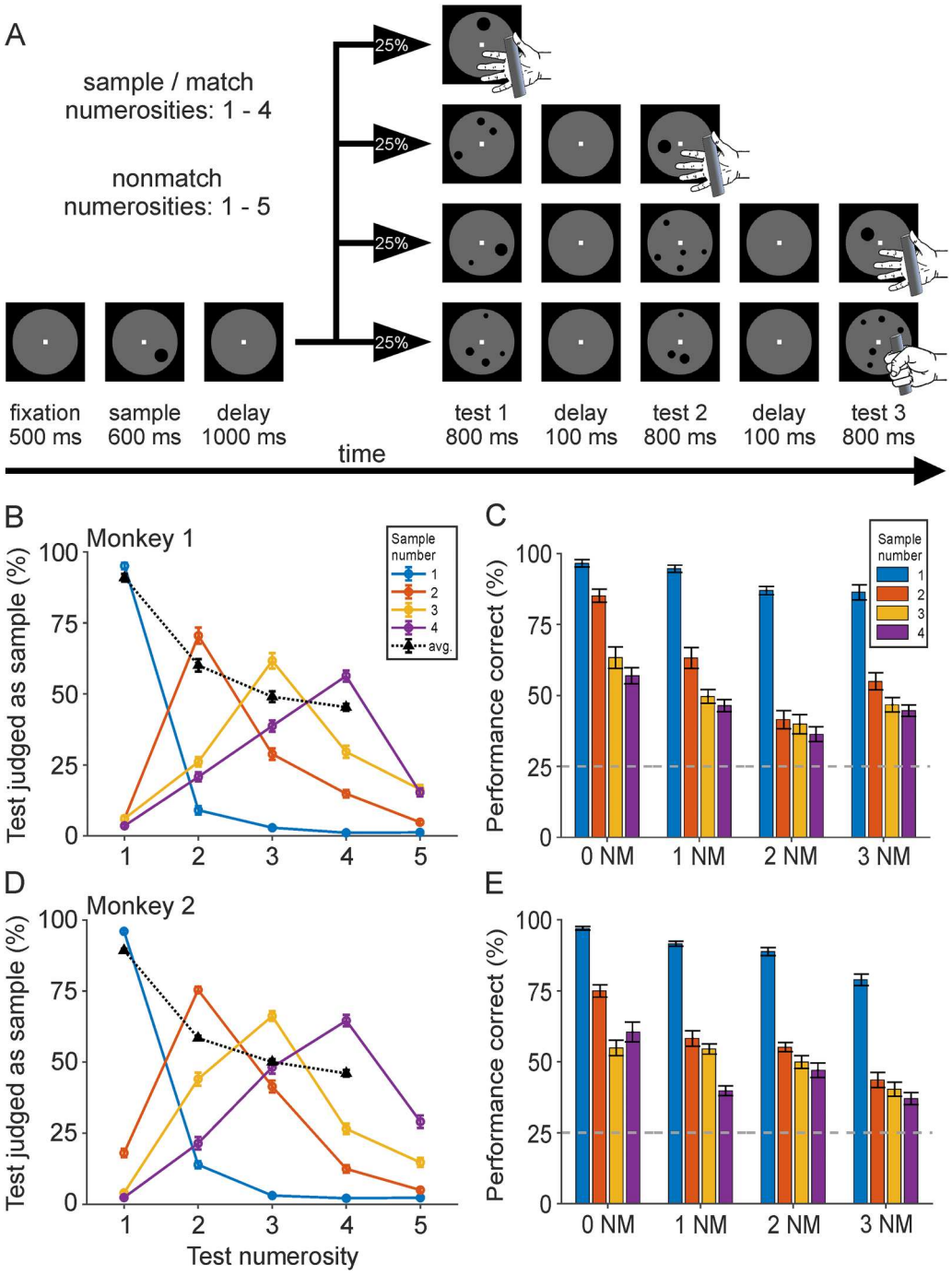


Fig. 1. Task protocol and behavioral performance. (A) Monkeys performed a delayed match-to-sample task with dot numerosities. A trial started when the monkey grabbed the response bar and fixated the fixation target; both actions were required to be maintained throughout the task until the monkey responded by releasing the bar. Numerosities ranging from 1 to 4 were presented during a sample phase that was followed by a delay period. The monkey had to respond whenever a numerosity that matched the sample was presented in one of the three subsequent Test phases. The monkeys had to refrain from responding for as long as non-matching numerosities (ranging from 1 to 5) were presented in the Test phases. Up to three sequential Test-phases (Test1-Test3) were presented pseudo-randomly in different trial conditions; in each Test-phase, a match or nonmatch-numerosity could appear. If the last Test3 period showed a nonmatch, the monkey was required to maintain the bar until display-offset. This resulted in a total four different trial conditions. Sample numerosities as well as trial conditions were presented in a balanced way. (B, D) Behavioral performance functions of the two monkeys to sample numerosities 1–4 (monkey 1: 27 sessions; monkey 2: 33 sessions). The functions reflect the probability that a monkey judged displays in the test periods as containing the same number of items as the sample numerosity (indicated in four colors, solid line). The peak data points of each colored function indicate correct sample matching. Data points in the flanks reflect error responses to nonmatch numerosities. Black dotted line indicates averaged performance for each sample numerosity. Error bars (very small) indicate the SEM across sessions. (C, E) Average correct performance of the two monkeys to the four sample numerosities and the four Test conditions (NM = nonmatch). Grey dashed line indicates chance level at 25 %. Error bars indicate the SEM across sessions.

2.2. General neuronal response properties

We recorded a total of 372 neurons from area VIP in the fundus of the intraparietal sulcus (IPS) and 509 neurons from the dorsolateral prefrontal cortex (dlPFC, centered around the principal sulcus) from two monkeys (Fig. 2A). The proportions of neurons for each monkey can be seen in Table S1. Throughout the paper, we focused our analyses on activity during the presentation of sample and test stimuli, including matching test stimuli to which the monkeys responded quickly by releasing a bar.

To determine the starting point of the neuronal analysis window, we utilized the neurons' visual response latencies. Visual response latencies were measurable in 51 % of the VIP neurons (190/372) and in 64 % of the PFC neurons (324/509). VIP neurons exhibited significantly shorter latencies compared to PFC neurons (median VIP: 111 ms, median PFC: 124 ms, Wilcoxon rank sum test, $Z = 2.6619$, $p = 0.0078$, two-sided) (Fig. 2B). We standardized the analysis window by using the first 25th percentile of the VIP (86 ms) and PFC latency distributions (100 ms) as the standard latency for all neurons, including those for which we could not determine individual latencies based on our criteria.

To ensure that neuronal responses were not contaminated by motor activity — that is, activity related to neurons involved in preparing and executing the lever release movement — we used the monkeys' reaction time (RT) distributions of correct responses to the matching test stimulus as endpoints of the duration of the neuronal analysis windows. In 84 % (monkey 1) and 89 % (monkey 2) of the cases, respectively, the monkeys responded after 400 ms or longer to the match. Therefore, we used 400 ms as the analysis window endpoint to ensure that neuronal activity is analyzed in equal time windows between the two monkeys and analyzed all correct trials in which the monkeys responded with an RT of at least 400 ms (trials with shorter RTs were excluded). Thus, VIP neurons were analyzed in a time window from 86 ms to 400 ms after sample/test onset, and PFC neurons from 100 ms to 400 ms. Qualitatively similar results in numerosity tuning were achieved by using 300 ms as the endpoint (cutoff) of the analysis window (see Fig. S2), confirming that the results were not due to motor-preparatory activity.

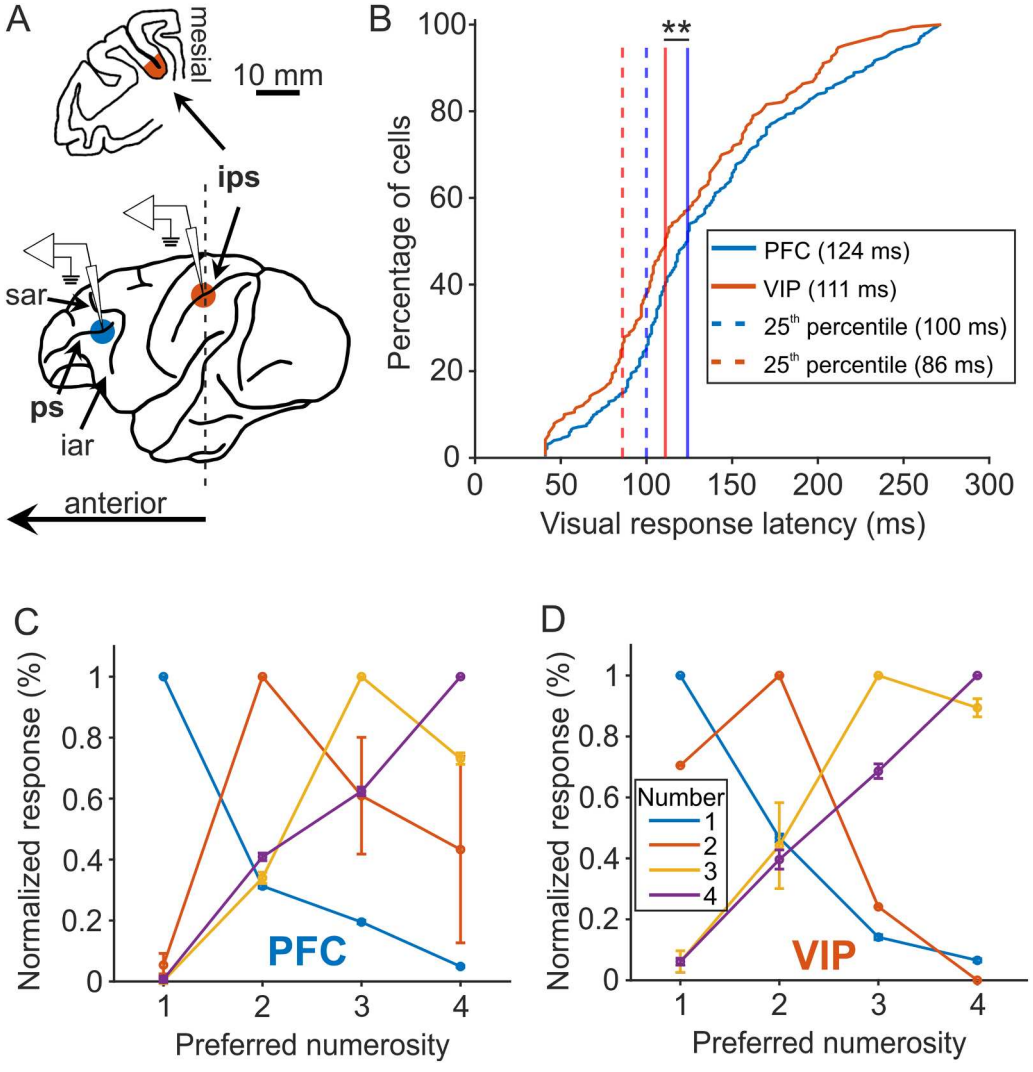


Fig. 2. Recording sites, visual response latencies and population tuning curves of PFC and VIP neurons. (A) Schematic diagram of the macaque brain illustrating the locations in PFC (blue) and VIP (red) where recordings were performed. Abbreviations: ips, intraparietal sulcus; ps, principal sulcus; sar, superior arcuate sulcus; iar, inferior arcuate sulcus. (B) Cumulative latency distributions across all neurons recorded in PFC (N = 324, blue) and VIP (N = 190, red) that showed significant firing rate change compared to baseline activity after sample onset. Numbers next to labels indicate the median latencies and latencies at the 25th percentiles in PFC and VIP. The vertical solid lines indicate the medians and the vertical dashed lines show the 25th percentiles. **$p < 0.01$. (C) Population tuning curves for numerosity-selective neurons in PFC (n = 68) during the sample period. Population tuning curves were obtained by averaging the normalized tuning curves of all single units preferring the same sample numerosity (see color-code). Error bars indicate the SEM. Numerosity 1, 2, 3, 4: $n = 35, 2, 11, 20$. (D) Population tuning curves for numerosity-selective neurons in VIP (n = 34) during the sample period. Same layout as in (C). Numerosity 1, 2, 3, 4: $n = 19, 1, 3, 11$.

2.3. Numerosity selective neurons

Using this time window, we initially identified neurons that exhibited numerosity selectivity (i.e., tuning to numerosity) during sample presentation. Since we were specifically interested in tuning changes over time, we only selected neurons with at least 18 correct trials per sample numerosity and 3 correct trials per Test numerosity (Test 1–3 periods) for further analysis. Thus, the analyzed population comprised a total of 369 PFC neurons and 254 VIP neurons that met all inclusion criteria.

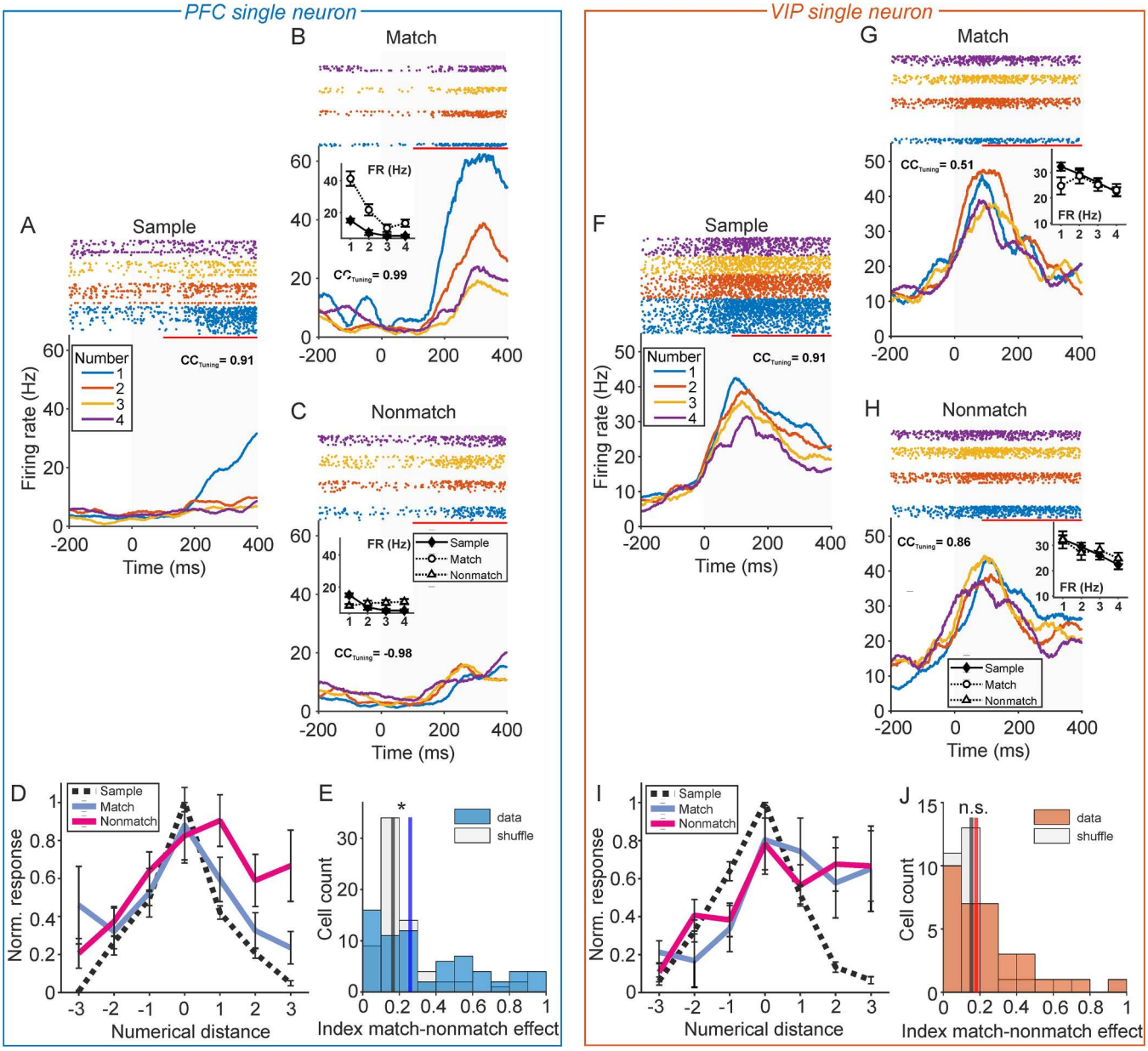


Fig. 3. Numerosity selectivity of single neurons during the sample and Test1 match and nonmatch periods. (A-C) Responses of an exemplary PFC neuron preferring numerosity 1 in the sample period (A), as match (B), and as nonmatch numerosities (c) during Test1. Top panels show dot-raster histograms (each line represents a trial; each dot represents an action potential) and bottom panels show smoothed spike-density functions (activity averaged and smoothed with a 75-ms Gaussian kernel). The colors of dots and spike density functions correspond to the numerosity of the sample/test stimulus. Time 0 is the onset of the stimulus in the sample (A) and Test1 period (B,C), indicated by the gray shaded areas. Insets show tuning curves calculated during the analysis window indicated by the red bar plotted at the top of the spike-density functions. The tuning function for the sample period (solid line) is plotted together with the tuning functions of match situation (circle, dotted line) in (B), or nonmatch situation (triangle, dotted line) in (c). CC_{Tuning}-values in spike-density histograms represent the cross-correlation coefficients between cross-validated tuning curves during the sample (A), a comparison between sample and match tuning curves (b), and a comparison between sample and nonmatch tuning curves (c). Error bars show the SEM. (D) Normalized and averaged numerosity tuning curves of all numerosity-selective neurons in PFC (N = 68) during the sample, match, and nonmatch phases. Prior to averaging, neuronal activity for each neuron was aligned to the preferred numerosity at center value 0 and plotted as a function of numerical distance. (E) Distribution of indices quantifying the differences between responses to match and nonmatch numerosities during Test1 in PFC (N = 68). The median of the real data (blue) indicated by the colored vertical line were tested against a shuffled distribution (grey) indicated by a black line. *$p < 0.05$. (F-H) Responses of an exemplary VIP neuron preferring numerosity 1 in the sample period (F) as match (G), and as nonmatch numerosities (H) during Test1. Same layout as in (A-C). (I) Normalized and averaged numerosity tuning curves of all numerosity-selective neurons in VIP (N = 34) during the sample, match, and nonmatch phases. Same layout as in (D). (J) Distribution of indices showing the strength of the match-nonmatch effect in VIP (N = 34). Real data (orange) were tested against a shuffled distribution (grey). n.s. not significant. Layout as in (E).

To identify numerosity-selective neurons, we applied a two-way ANOVA with main factors of 'numerosity' and 'protocol', evaluated at $p < 0.01$. We only considered neurons that were exclusively numerosity-selective, showing a main effect for numerosity but no main effect for protocol or interaction. By combining both selection criteria, we found 34 numerosity-selective neurons with a sufficient number of trials in VIP (34/254; 13.4 % of the recorded neurons) and 68 neurons in PFC (68/369; 18.4 % of the recorded neurons). In PFC, only 3.7 % of all neurons were non-exclusivelynumerosity selective, i.e. they showed in addition to a main effect 'numerosity', a main effect 'protocol' and/or an interaction between 'numerosity' and 'protocol'. In VIP, only 2.2 % of all neurons were non-exclusively numerosity selective. The proportions of exclusively numerosity-selective neurons reported in the current study are comparable to what we found in previous studies with other monkeys (on average, VIP: 11 %; PFC: 19 %) (Diester and Nieder, 2007; Viswanathan and Nieder, 2015; 2020). The average tuning functions for selective neurons in the two regions are shown in Fig. 2C,D.

To assess the robustness of numerosity tuning, we cross-validated the tuning functions of individual selective neurons by correlating tuning curves derived from odd versus even trials. In both brain areas, we found cross-correlation coefficients close to 1 (PFC: $CC_{median} = 0.89$, VIP: $CC_{median} = 0.90$), indicating strong and stable numerosity tuning in both PFC and VIP. The correlation coefficients did not differ significantly between PFC and VIP (Wilcoxon rank-sum test, $Z = -0.0674$, $p = 0.9462$, two-sided), suggesting that tuning stability was comparable between neurons in both regions (Suppl. Fig. S3).

We next evaluated the numerosity tuning of sample-selective neurons during the delay working memory period. Consistent with previous findings (Nieder et al., 2002), a similar proportion of sample-selective numerosity neurons were also tuned to numerosity during the first delay period after the sample in both the PFC and VIP (PFC: 25/68 = 37 %; VIP: 15/34 = 44 %). Moreover, numerosity tuning between the sample and the first delay period was correlated in PFC neurons but not in VIP neurons (Figure S4).

After calculating neuronal tuning functions for each of these neurons, we explored the stability of their tuning relative to sample presentation in two dimensions. First, tuning stability was examined at the identical trial time when Test1 was either a match or a nonmatch. We hypothesized that tuning remained stable to different degrees in VIP and PFC when the monkey was presented with the matching numerosity it was searching for, but deteriorated when the monkey saw a deviating numerosity for what it was not seeking in a given trial. Second, tuning stability was investigated across all three succeeding test periods (Test1 to Test3) when the respective test periods were a match. We anticipated that tuning to the matching numerosity would remain stable across time when the monkey was searching for it throughout the trial, potentially with differences between VIP and PFC. We start with the first question: tuning as a function of matching or nonmatching numerosity.

2.4. Tuning stability to match and nonmatch numerosity in the Test1 period

We initially examined the tuning functions of single neurons. Fig. 3A-C illustrates the detailed discharges and the tuning functions (insets) of an example PFC neuron tuned to numerosity 1 during sample presentation for the match and nonmatch Test1 period. When a matching numerosity was displayed, this neuron's tuning remained stable and exhibited enhanced firing rates for the match period (Fig. 3B). In contrast, this neuron's numerosity selectivity was significantly diminished when a nonmatch numerosity was presented in the same Test1 period, resulting in a more or less flat tuning function (Fig. 3C). Similar stability in tuning to match numerosities, but degraded tuning for nonmatch numerosities, was observed for the majority of PFC neurons.

In contrast, VIP neurons displayed less stable tuning. Fig. 3F-H portrays an example VIP neuron tuned to numerosity 1. Despite clear tuning to the sample numerosities (Fig. 3F), this neuron exhibited poorer tuning to the match numerosity (Fig. 3G) and similarly deteriorated tuning to the nonmatching numerosities (Fig. 3H).

To quantify tuning stability as a function of match versus nonmatch numerosity, we calculated normalized tuning curves with firing rates during sample presentation as the reference for all stimulus presentation phases. Subsequently, the normalized tuning functions were averaged according to the respective preferred sample numerosities for match and nonmatch conditions. Finally, the normalized tuning functions for match and nonmatch conditions were aligned relative to the preferred sample numerosities and plotted as a function of numerical distance from the respective preferred numerosities, with the preferred numerosities set to numerical distance 0. This approach allowed us to assess how neuronal tuning varied across different stimulus conditions and numerical distances from the preferred numerosities.

In PFC, the population-tuning curve for the match condition was comparable to the sample tuning curve and showed a corresponding preferred numerosities (Fig. 3D). Interestingly, this was not the case for the nonmatch presentation; here, the population-tuning curve was asymmetric, and the peak deviated from the preferred numerosity. We have previously shown that numerosity tuning curves are asymmetric on a linear number scale (e.g., Nieder and Miller, 2003), with this effect being less visible in sharp tuning functions during the sample period and more pronounced as neurons become less selective in the Test1 period. To quantify this difference between sample and Test1 periods, we fitted Gauss functions to the individual tuning curves and derived the sigma-value as a measure of tuning width, or tuning selectivity, respectively.

The Gauss fits described the tuning functions well. In both PFC and VIP, the goodness-of-fit (R^2) for the fits during sample period (the reference period) were close to 1 (PFC: median $R^2 = 0.95$; VIP: median $R^2 = 0.91$). The R^2 values were higher during the sample period compared to the Test periods, but no difference was found between the match (PFC: median $R^2 = 0.71$; VIP: median $R^2 = 0.57$) and nonmatch periods (PFC: median $R^2 = 0.78$; VIP: median $R^2 = 0.65$), neither in PFC (Wilcoxon signed-rank test, $Z = -0.5214$, $p = 0.6021$, two-sided) nor in VIP (Wilcoxon signed-rank test, $Z = -0.9704$, $p = 0.3318$, two-sided).

In PFC, no difference in tuning width was found for sample and match conditions (median sample: $\sigma = 1.24$, median match: $\sigma = 1.45$, Wilcoxon signed rank test, $Z = -1.8392$, $p = 0.1977$, two-sided, with Bonferroni correction), indicating that PFC neurons maintained their numerosity selectivity between these two trial phases. In contrast, the tuning functions in the nonmatch condition were significantly wider, both compared to the match condition (median match: $\sigma = 1.45$, median nonmatch: $\sigma = 2.01$, Wilcoxon signed rank test, $Z = -2.4686$, $p = 0.0407$, two-sided, with Bonferroni correction, $n = 68$) and the sample period (median sample: $\sigma = 1.23$, median nonmatch: $\sigma = 2.01$, Wilcoxon signed rank test, $Z = -4.3261$, $p < 0.001$, two-sided, with Bonferroni correction). Thus, tuning selectivity of PFC neurons specifically decreased if a nonmatch numerosity was shown.

In VIP, in contrast, the population-tuning curves for both the match and nonmatch conditions were deteriorated relative to the sample tuning curve (Fig. 3I). Gauss function fitted to the tuning curves showed that the tuning widths for both the match condition (median sample: $\sigma = 1.42$, median match: $\sigma = 2.35$, Wilcoxon signed rank test, $Z = -2.9491$, $p = 0.0096$, two-sided, with Bonferroni correction) and the nonmatch condition (median sample: $\sigma = 1.42$, median nonmatch: $\sigma = 2.58$, Wilcoxon signed rank test, $Z = -3.1372$, $p = 0.0051$, two-sided, with Bonferroni correction) were significantly larger compared to the sample period. In addition, no tuning widths difference was observed between match and nonmatch conditions (median match: $\sigma = 2.34$, median nonmatch: $\sigma = 2.58$, Wilcoxon signed rank test, $Z = 0.6240$, $p = 1$, two-sided, with Bonferroni correction, $n = 34$). This speaks for a deterioration of tuning selectivity in VIP from sample to Test1 that is independent of whether a match or a nonmatch numerosity is presented.

We quantified the magnitude of firing rate differences to the preferred numerosity between match and nonmatch conditions in both

brain areas by calculating an index capturing the match-versus-nonmatch effect (Miller et al., 1996; see Methods). The higher the index, the stronger the effect (values of 0 indicate equal responses to preferred numerosities in matches and nonmatches). We compared the medians of each distribution within each brain area with a shuffled distribution representing chance values. In PFC, the real index (median 0.26) was significantly higher than the shuffled index (0.17) ($Z = 0.9567$, $p = 0.0209$, Wilcoxon signed rank test, $n = 68$) (Fig. 3E). In VIP, in contrast, the real index (median 0.18) was not different from the shuffled index (0.15) ($Z = 0.9935$, $p = 0.3387$, Wilcoxon signed rank test, $n = 34$) (Fig. 3**J**). Moreover, a direct comparison between brain areas showed that the real indices for PFC were significantly larger compared to VIP ($Z = 1.6824$, $p = 0.0462$, Wilcoxon rank sum test). In sum, match-nonmatch effects were prominent in PFC, but absent in VIP.

A matching numerosity in the Test1 period could either enhance the firing rates to the preferred numerosity, leading to a positive index, or suppress the firing rates to the preferred numerosity, causing a negative index (Miller et al., 1996). We therefore categorized the neurons into match-enhancement and match-suppression cells and compared the indices for the two separate classes in PFC and VIP. Negative indices of match-suppression cells were rectified by using the absolute values for better comparison, thus absolute index values larger than 0 indicate differential responses to preferred numerosities in matches and nonmatches.

The indices for neurons exhibiting a match-enhancement effect in the PFC ($n = 31$) and VIP ($n = 14$), respectively, are shown in Figs. 4A and 4B. The median enhancement index for the PFC was 0.23, compared to 0.15 in the VIP. We compared these medians within each brain area using a shuffled distribution representing chance values. In the PFC, the real index (median 0.23) was significantly higher than the shuffled index (0.13) ($Z = 2.4785$, $p = 0.0132$, Wilcoxon signed-rank test, $n = 31$). In contrast, the real index in the VIP (median 0.15) was not significantly different from the shuffled index (0.12) ($Z = 1.5380$, $p = 0.1240$, Wilcoxon signed-rank test, $n = 14$). Furthermore, direct comparisons between brain areas showed that the real indices for the PFC tended to be larger than those for the VIP ($Z = 1.2871$, $p = 0.0990$, Wilcoxon rank-sum test).

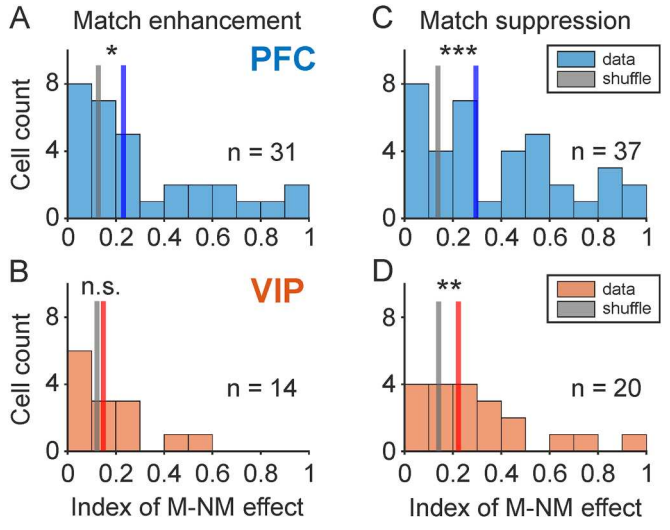


Fig. 4. Distribution of match-nonmatch indices showing the strength of the match enhancement effect and match suppression effect in PFC and VIP in the Test1 period. (A) Indices for neurons exhibiting a match enhancement effect in the PFC. The index is calculated as the absolute value of the difference between match and non-match responses, divided by their sum. The median of the real data, indicated by the blue vertical line, was tested against a shuffled distribution, whose median is indicated by a black line. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$; n.s. not significant. (B) Indices for neurons exhibiting a match enhancement effect in VIP. Layout as in (A). (C) Indices for neurons exhibiting a match suppression effect in PFC. Layout as in (A). (D) Indices for neurons exhibiting a match suppression effect in VIP. Layout as in (A).

The indices for neurons showing a match-suppression effect in the PFC ($n = 37$) and VIP ($n = 20$) are depicted in Figs. 4C and 4D. In both the PFC and VIP, the real indices (median 0.30 and median 0.22) were significantly higher than the shuffled indices (0.14 for both) ($Z = 3.6199$, $p < 0.001$, Wilcoxon signed-rank test, $n = 37$ for PFC; $Z = 2.6157$, $p = 0.0089$, Wilcoxon signed-rank test, $n = 20$ for VIP). Furthermore, we found no significant difference when comparing the distributions between the two brain areas ($Z = 1.1455$, $p = 0.1260$, Wilcoxon rank-sum test). In summary, significant match enhancement effects were found only in the PFC, while significant match suppression effects occurred in both the PFC and VIP with similar magnitudes.

Finally, to explore the information contained in all PFC and VIP neurons, regardless of their numerosity tuning, we conducted a time-resolved percent explained variance (PEV, ω^2) analysis (PFC: $n = 369$, VIP: $n = 254$; see Methods for details). Percent explained variance (PEV) is a statistical measure that quantifies how much of the variability in our neuronal dataset is explained by the factor 'match' or 'nonmatch'. Both PFC and VIP neurons represented match information throughout the sample and delay periods (greater than the significance threshold of the shuffled distribution at $p < 0.01$), while nonmatch information was still absent ($p < 0.01$) (Fig. 5). Nonmatch information was only represented by neuron populations during the Test1 period, with important differences between PFC and VIP. In PFC, match information continued to increase and consistently remained higher than nonmatch information ($Z = 3.8263$, $p < 0.001$, Wilcoxon signed-rank test, $n = 369$). In VIP, however, match information remained constant during Test1, showing no significant difference from nonmatch information ($Z = 1.6477$, $p = 0.0994$, Wilcoxon signed-rank test, $n = 254$). This suggests that PFC neurons, but not VIP neurons, irrespective of numerosity tuning, discriminate between match and nonmatch information. These effects were also observed in PFC and VIP neurons of individual monkeys. In monkey 2, however, the match information seems to be more strongly maintained in the population code during the delay and does not appear to be updated in VIP, whereas the non-match information is (Suppl. Fig. S5).

2.5. Population decoding of match and nonmatch numerosity in the Test1 period

So far, our analysis of single-neuron responses suggests that numerosity-selective PFC neurons have a privileged role in coding and maintaining sought after numerical information. We further tested this by a population decoding analysis using a support vector machine (SVM) classifier. SVM classifiers offer a powerful approach to decoding neural activity by simultaneously considering the combined activity of multiple neurons, capturing complex multidimensional patterns and non-linear relationships underlying population coding. This makes classifier decoding more robust to noise and complexity than simple averaging approaches. We trained the classifier on the firing rates of numerosity-selective neurons (balanced neuron numbers for the two brain areas) during the sample period. Next, we tested the trained classifier on its accuracy to discriminate numerosities during the match or nonmatch Test1 period (6-fold cross validated, 100 times resampled; see Methods).

Fig. 6A depicts the mean decoding accuracy for the sample (as a reference), the match Test1 period, and the nonmatch Test1 period in both PFC and VIP, respectively. In all conditions and brain areas, the decoding was significantly above chance (larger than the 95th percentile of the respective shuffled distribution; PFC: Sample, Match, Nonmatch [36.11 %, 33.33 %, 31.94 %]; VIP: Sample, Match, Nonmatch [35.42 %, 31.25 %, 32.64 %]).

We next compared the classifier's decoding quality between PFC and VIP neurons. In the sample period, decoding accuracy was significantly higher in PFC (mean 60 ± 0.66 %) than in VIP (53 ± 0.66 %) (permutation test, $p < 0.001$, with alpha correction, $n = 100$). As expected,

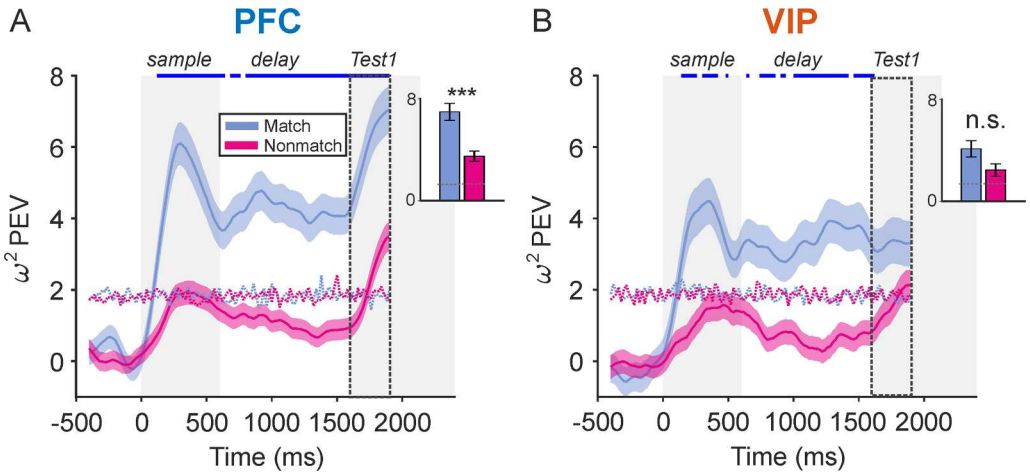


Fig. 5. Sliding-window ω^2 PEV from sample to Test1 period quantifying the encoding of match (same as sample) and nonmatch (different from sample) numerosity. Sliding-window ω^2 PEV was computed for all neurons recorded with at least 18 correct trials per sample numerosity and 3 correct trials per test numerosity in each test phase. (A) PFC neurons (n = 369), (B) VIP neurons (n = 254). Bars above the curves represent time bins where match information exceeded nonmatch information (blue: $p < 0.05$). *Insets*: Mean ω^2 PEV across the Test1 period for match and nonmatch conditions (derived from inside dashed rectangle in the main figure). Dotted lines indicate the significance threshold ($p < 0.01$, one-sided) from shuffled trial labels. Error bars and shaded regions indicate SEM across neurons; ***$p < 0.001$; n.s., not significant.

accuracy dropped in both brain areas for the subsequent Test1 periods. Averaged decoding accuracy in the match condition dropped to 42 % (± 0.81) in PFC, and 37 % (± 0.64) in VIP. Still, classifier accuracy in the match condition remained significantly higher for PFC neurons compared to VIP neurons (permutation test, $p < 0.001$, with alpha correction)

Important differences in decoding emerged for match and nonmatch conditions within each brain area. In PFC, the decoding accuracy for the match period (42 ± 0.81 %) was significantly higher than in the nonmatch period (35 ± 0.64 %) (permutation test, $p < 0.001$, with alpha correction). This was not the case for VIP with a similarly low decoding performance for match and nonmatch conditions (match: 37 ± 0.64 %, nonmatch: 38 ± 0.60 %, permutation test, $p = 0.283$, with alpha correction). A second classifier model, trained and tested with nearly double the number of trials, yielded qualitatively identical results (see Fig. S6). Overall, decoding performance of the classifier based on an equal number of numerosity-selective neurons shows that the PFC represents the sought after matching numerosity better than the nonmatching numerosity, and overall better than VIP neurons which did not differentiate between match and nonmatch numerosities.

Furthermore, we calculated the average performance of the classifier as a function of distance from the real numerosity and used, for each iteration, the sigma-value derived from Gauss fits as a measure of tuning width analogous to the population-tuning curves. Within both brain areas, the tuning functions during the sample period were significantly narrower compared to match and nonmatch conditions, respectively (sample PFC: median $\sigma = 1.01$, sample VIP: median $\sigma = 1.27$, $p < 0.001$). Importantly, and similar to the previous results with averaged tuning curves, the PFC classifiers' tuning functions in the nonmatch condition were significantly wider than those calculated for the match situation (Fig. 6B, median match: $\sigma = 1.86$, median nonmatch: $\sigma = 2.68$, Wilcoxon rank sum test, $Z = -7.2556$, $p < 0.001$, two-sided, with Bonferroni correction, $n = 100$). This was not the case for tuning functions of the VIP (Fig. 6C) for which the width in match and nonmatch conditions were not significantly different (median match: $\sigma = 2.50$, median nonmatch: $\sigma = 2.26$, Wilcoxon rank sum test, $Z = 2.0732$, $p = 0.1145$,

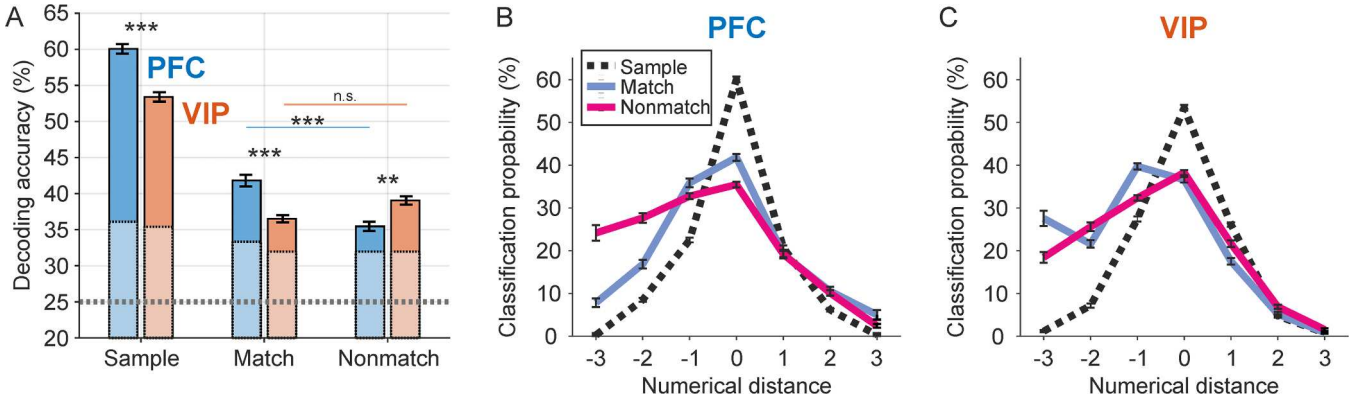


Fig. 6. Decoding numerosity with a SVM classifier from numerosity-selective neurons in the sample and match and nonmatch Test1 periods. (A) Mean decoding accuracy for sample, match, and nonmatch phases. The classifier was trained on sample activity and then tested on sample activity (odd versus event trials cross validation), match activity in Test1, and nonmatch activity in Test1. PFC and VIP cell counts were balanced by creating two pseudo populations of equal neuron numbers. Grey dashed line indicates chance level at 25 %. Dashed lines within the bars depicts significance threshold ($p < 0.05$, one-sided). Error bars show the SEM across 100 resamples. ***$p < 0.001$; **$p < 0.01$; n.s. not significant. (B) Normalized and averaged numerosity tuning curves derived from classifier performance on PFC neurons (N = 34) during the sample, match, and nonmatch phases. Classification probability for each resample is aligned to the preferred numerosity at center value 0 and plotted as a function of numerical distance. Error bars show the SEM across 100 resamples. (C) Normalized and averaged numerosity tuning curves derived from classifier performance on VIP neurons (N = 34) during the sample, match, and nonmatch phases. Layout as in (B).

two-sided, with Bonferroni correction, n = 100).

2.6. Tuning stability to matching numerosity across time and test periods

After comparing tuning to match and nonmatch, we switched to investigating tuning stability across time. To that aim, we explored tuning of single numerosity-selective neurons in PFC and VIP to matching numerosities along the three succeeding test periods (Test1 to Test3). Fig. 7A-D shows the detailed discharges and the tuning functions (insets) of an example PFC neuron tuned to sample numerosity 4 across stimulus repetitions. This PFC neuron's tuning remained stable for all three succeeding test phases until the matching numerosity was presented. Similar stability in tuning was seen for the majority of PFC neurons. In contrast, such a robust encoding was rarely observed in VIP neurons. Fig. 7E-F depicts an example VIP neuron tuned to numerosity 1. This neuron showed a deteriorated tuning with the presentation of more test numerosities and had lost tuning during test periods (Fig. 7G, H). Inspection of single neuron activity indicated that the PFC robustly encodes the target numerosity across repetitions, whereas VIP deteriorated with repetitions. This suggests that VIP neurons are primarily involved during the early stages of cortical processing, while PFC neurons maintain the target number across temporal delays during delayed searches.

To explore tuning stability as a function of test period repetition across neurons, we again computed normalized tuning curves (see description above) and plotted them aligned to the preferred numerosities of each neuron.

In PFC, population-tuning curves remained stable and showed rather symmetric peak curves in all three test periods following the sample presentation (Fig. 8A). Compared to the reference sample period, the preferred numerosity in PFC remained the same across the three subsequent test periods ($peak_{sample} = 1.00 \pm 0.00$, $peak_{test1} = 0.88 \pm 0.20$, $peak_{test2} = 0.79 \pm 0.14$, $peak_{test3} = 0.80 \pm 0.13$, permutation test, Sample/Test1: $p = 0.6795$, Sample/Test2: $p = 0.2957$, Sample/Test3: $p = 0.2672$, $n = 68$). Furthermore, we compared again the tuning width of the individual tuning curves. We found no difference between the sigma-values derived from the three test periods (median Test1: $\sigma = 1.45$, median Test2: $\sigma = 1.84$, median Test3: $\sigma = 1.79$, Wilcoxon signed rank test, Test1/Test2: $Z = -1.7842$, $p = 0.2975$, Test1/Test3: $Z = -1.9981$, $p = 0.1828$, Test2/Test3: $Z = -0.6416$, $p = 1$, with Bonferroni correction, $n = 68$).

In VIP, a different picture emerged (Fig. 8B). While VIP population tuning curves were still comparable for the second stimulus presentation (Test1) ($peak_{sample} = 1.00 \pm 0.00$, $peak_{test1} = 0.80 \pm 0.18$, permutation test, $p = 0.2971$, $n = 34$), both the amplitude and the shape of the tuning functions deteriorated with more stimulus repetitions. For the

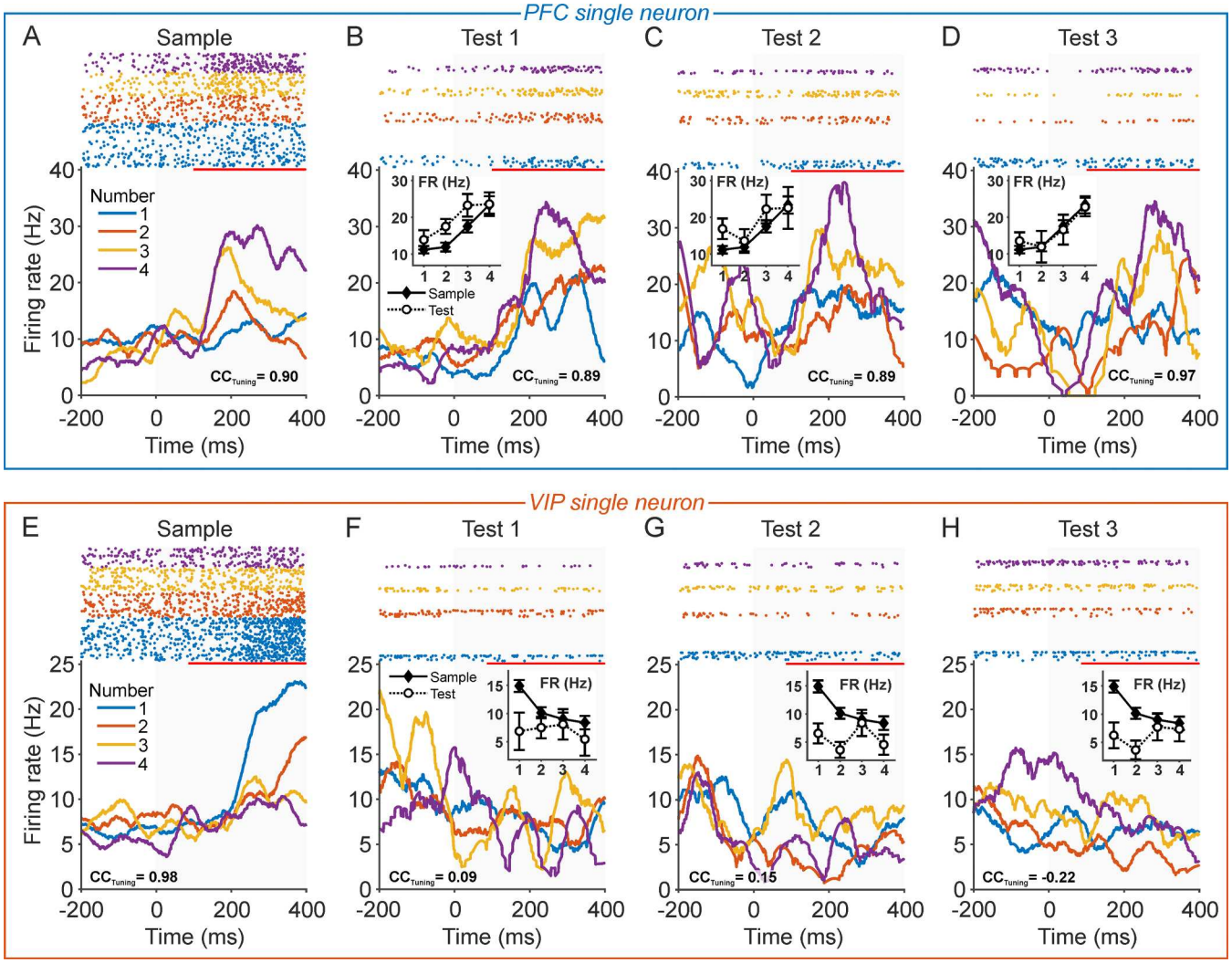


Fig. 7. Numerosity selectivity of single neurons during the sample and Test1–3 match periods. (A-D) Responses of an exemplary PFC neuron preferring numerosity 4 in the sample period (A) to match numerosities in Test1 (B), Test2 (C), and Test3 (D) periods. Same layout as in Fig. 3a-c. (E-H) Responses of an exemplary VIP neuron preferring numerosity 1 in the sample period (E) to match numerosities in Test1 (F), Test2 (G), and Test3 (H) periods. Same layout as in Fig. 3E-G.

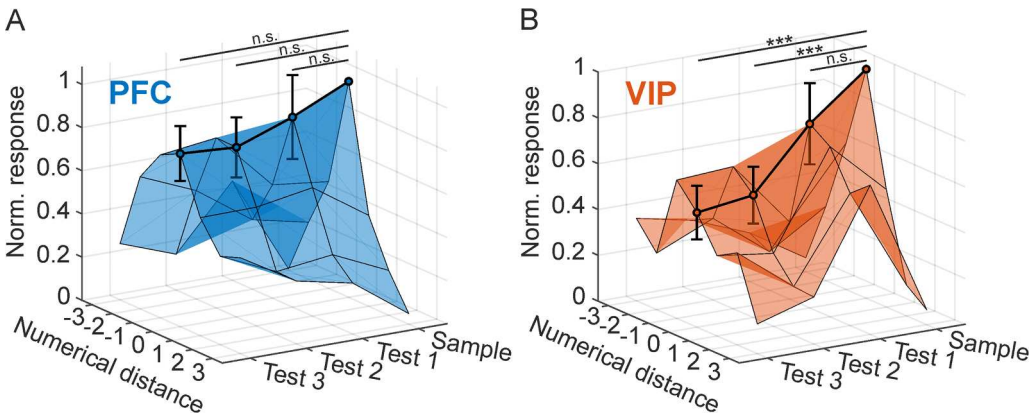


Fig. 8. Time course of tuning in numerosity-selective neurons across the sample and the three successive Test periods. (A) Normalized and averaged numerosity tuning curves of all numerosity-selective neurons in PFC (N = 68) during the sample, Test1, Test2, and Test3 phases mapped as a 3D-surface plot. Prior to averaging, neuronal activity for each neuron was aligned to the preferred numerosity at center value 0 and plotted as a function of numerical distance. Tuning remained stable across Test periods. Error bars indicate SEM across neurons. n.s. not significant. (B) Normalized and averaged numerosity tuning curves of all numerosity-selective neurons in VIP (N = 34) during the sample, Test1, Test2, and Test3 phases mapped as a 3D-surface plot. Same layout as in (A). Tuning deteriorated across Test periods. Error bars indicate SEM across neurons. ***p < 0.001; n.s. not significant.

Test2 and Test3 periods, the VIP population-tuning curves were attenuated, asymmetric, and the peaks were off-center. The normalized peak values in Test2 and Test3 were significantly different to the peak values of the sample period ($peak_{sample} = 1.00 \pm 0.00$, $peak_{test2} = 0.53 \pm 0.13$, $peak_{test3} = 0.50 \pm 0.12$, permutation test, Sample/Test1: $p < 0.001$, Sample/Test2: $p < 0.001$, $n = 34$). These results support the initial finding that PFC neurons maintain their tuning across multiple match presentations, whereas the numerosity tuning of VIP neurons degrades.

Next, we extended the index calculations introduced during the Test1 period (Fig. 3E,J) to the Test2 and Test3 periods (Fig. 9). Besides comparing the neuronal response between the match and nonmatch of the Test1 period, we investigated also how the indices are distributed when using the neuronal response during the other match periods (match Test2 vs. nonmatch Test1; match Test3 vs. nonmatch Test1).

In PFC, we found that real indices to all Test periods (median Test1: 0.26; Test2: 0.37; Test3: 0.30) were significantly higher compared to the shuffled chance values (0.17, 0.17, 0.18, respectively) (match Test1 vs. nonmatch Test1: $Z = 0.9567$, $p = 0.0209$; match Test2 vs. nonmatch Test1: $Z = 3.6973$, $p < 0.001$; match Test3 vs. nonmatch Test1: $Z = 2.6333$, $p = 0.0085$; Wilcoxon signed rank test, $n = 68$) (Fig. 9A,C,E).

In VIP, this was not the case. The real indices to all Test periods (median Test1: 0.18; Test2: 0.14; Test3: 0.18) were not significantly higher compared to the shuffled chance values (0.15, 0.14, 0.15, respectively) (match Test1 vs. nonmatch Test1: $Z = 0.9935$, $p = 0.3387$; match Test2 vs. nonmatch Test1: $Z = 0.5213$, $p = 0.6022$; match Test3 vs. nonmatch Test1: $Z = 1.5087$, $p = 0.1314$; Wilcoxon signed rank test, $n = 34$) (Fig. 9B,D,F).

Furthermore, the indices were significantly higher in PFC compared to VIP for the first two Test periods (match Test1 vs. nonmatch Test1:

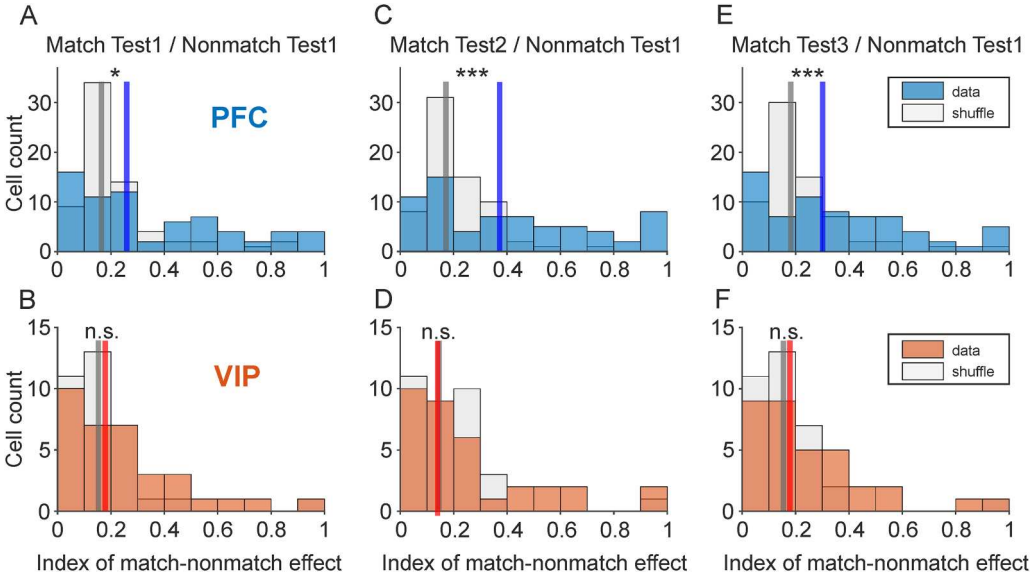


Fig. 9. Indices quantifying the differences between responses to Test1–3 match and Test1 nonmatch numerosities. (A,C,E) Distribution of indices showing the strength of the match-nonmatch effects across the three Test periods relative to the nonmatch responses in the Test1 period in PFC (N = 68). The distribution of indices is shown for the Test1 (A, same as in Fig. 3*E*), Test2 (c), and Test3 periods (E). The median of the real data indicated by the colored vertical line in each Test period was tested against a shuffled distribution indicated by a black line. ***p < 0.001; **p < 0.01; *p < 0.05; n.s. not significant. (B,D,F) Distribution of indices showing the strength of the match-nonmatch effects across the three Test periods relative to the nonmatch responses in the Test1 period in VIP (N = 34). The distribution of indices is shown for the Test1 (B, same as in Fig. 3*J*), Test2 (D), and Test3 periods (F). Same layout as in (A,C,E).

Z = 1.6824, p = 0.0462; match Test2 vs. nonmatch Test1: Z = 2.7652, p = 0.0028), and tended to be larger in the last Test period (match Test3 vs. nonmatch Test1: Z = 1.5335, p = 0.0626; Wilcoxon rank sum test). Thus, in PFC, higher tuning indices indicated more pronounced differences between match and nonmatch across all three test periods, an effect that was absent in VIP.

We compared the tuning properties in more detail by cross-correlating the entire tuning curves of the sample period (as a reference) with those of the subsequent test periods to derive the cross-correlation coefficients (CC_{Tuning}) (Diester and Nieder, 2007, see Material and Methods). The higher the coefficients, the more similar are the tuning functions in sample and in the respective test phases, i.e. a CC_{Tuning} of 1 would indicate identical tuning profiles, whereas a CC_{Tuning} of −1 would signify tuning functions exactly inverted relative to each other. As ground truth, we also determined the CC_{Tuning} for the sample itself by comparing the tuning functions derived from odd trial numbers with that derived from even trial numbers.

Example data can be seen for individual neurons in Fig. 7; here, the example PFC neuron shows a stable tuning across the test phases (Fig. 7A-D). The tuning functions are highly similar (i.e., overlapping), resulting in a high CC_{Tuning} near 1. In contrast, the example VIP neuron shows deteriorated tuning with increasing test phase repetitions, resulting in close-to-zero CC_{Tuning} (Fig. 7E-H).

In PFC, the CC_{Tuning} of all numerosity-selective neurons in all three test periods were significantly higher than zero ($median_{test1}$ = 0.7091, $median_{test2}$ = 0.5518, $median_{test3}$ = 0.5285; bootstrapping the medians, 95th confidence intervals Test1 [0.38, 0.82], Test2 [0.27, 0.70], Test3 [0.30, 0.74]) (Fig. 10B-D). This again suggests that the initial sample tuning remains stable over the three test phases.

In VIP, in contrast, the CC_{Tuning} in numerosity-selective VIP neurons showed overall smaller coefficient values in the test periods ($median_{test1}$ = 0.5602, $median_{test2}$ = 0.2846, $median_{test3}$ = 0.3310, Fig. 10F-H). The median CC_{Tuning} for the Test1 and Test2 periods were still significantly higher than zero (bootstrapping the medians, 95th confidence intervals Test1 [0.13, 0.74], Test2 [0.01, 0.67]), whereas the median for the Test3 period were not significantly different from zero (Fig. 10H; 95th confidence intervals Test3 [-0.09, 0.67]). Thus, the neurons entire tuning functions were more robustly encoded across the test periods in PFC neurons than in VIP neurons.

2.7. Population decoding of matching numerosity across time and test periods

Finally, we applied a population decoding analysis as previously used for the match-nonmatch comparison using a support vector machine (SVM) classifier. We trained the classifier on the firing rates of (area-balanced) numerosity-selective neurons during the sample period and tested the trained classifier during the Test1–3 periods again only for match conditions. Fig. 11 shows the decoding accuracy for the sample period and the three test periods for both brain areas. As expected, accuracy dropped in both brain areas for the subsequent test periods, but interesting differences between PFC and VIP emerged.

In PFC, average decoding accuracy dropped from 60 % (± 0.66) during sample presentation to 42 % (± 0.81), 43 % (± 0.70), and 40 % (± 0.73) during Test1, Test2, and Test3, respectively (Fig. 11A). In all cases, the classifier performances were significantly higher than chance level (larger than the 95th percentile of the shuffled distribution [33.33 %, 31.94 %, 33.33 %]) and stable across the test periods. The decoding accuracies in PFC between the Test1 and Test2, as well as between Test1 and Test3 were not significantly different (permutation test, Test1/Test2: p = 1, Test1/Test3: p = 0.396, with alpha correction), whereas the decoding success for Test2 period was slightly better than Test3 (permutation test, p = 0.01, with alpha correction).

In VIP, average decoding accuracy decreased from 53 % (± 0.66) during sample presentation to 37 % (± 0.70), 37 % (± 0.71) and 32 % (± 0.64) during Test1, Test2, and Test3, respectively (Fig. 11A). While accuracy was still above chance during Test1 and Test2 (larger than the 95th percentile of the shuffled distribution [31.94 %, 33.33 %]), accuracy dropped to chance level for Test3 (smaller than the 95th percentile of the shuffled distribution [34.03 %]). Decoding performance in VIP for Test1 and Test2 were comparable (permutation test, p = 1, with alpha correction) but relative to those dropped for the Test3 period (permutation test, Test1/Test3: p < 0.001, Test2/Test3: p < 0.001, with alpha correction).

When comparing PFC and VIP, we found that decoding accuracy in PFC was significantly higher in all three test periods (permutation test,

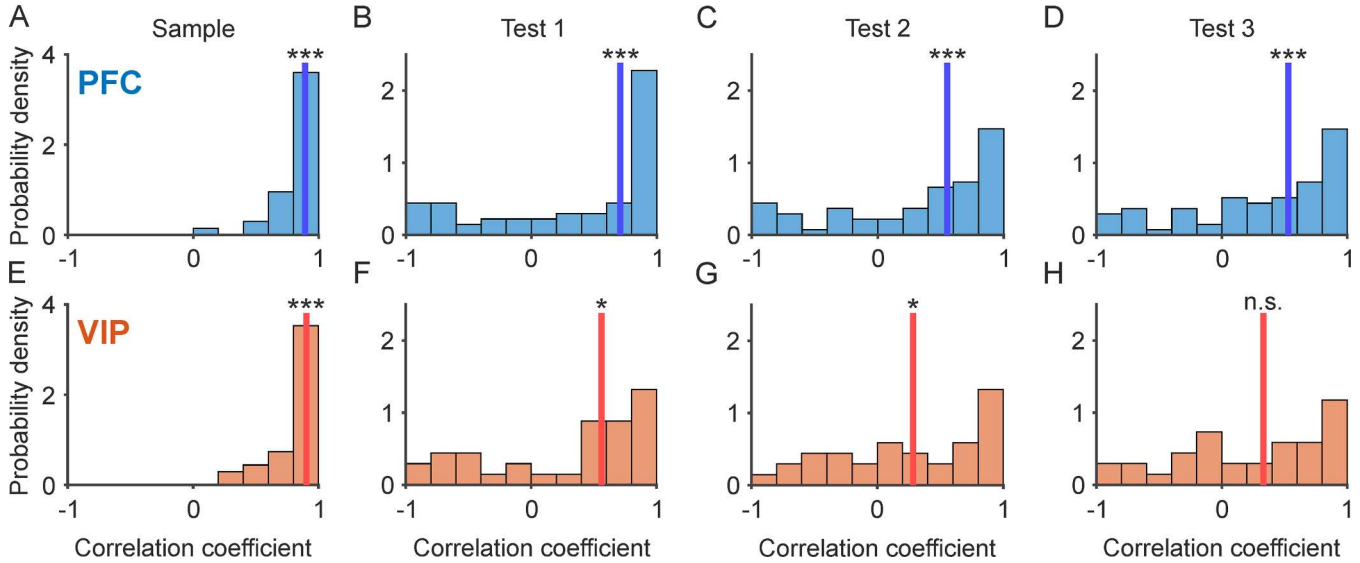


Fig. 10. Comparison of numerosity tuning based on tuning curve cross-correlations. (A-D) Probability density functions of cross-correlation coefficients (CC_{Tuning}) calculated for comparisons of tuning curves of numerosity selective PFC neurons (N = 68) during the sample (cross validation of odd versus even trials) (A), match Test1 (sample versus match Test1) (B), match Test2 (sample versus match Test2) (C), and match Test3 period (sample versus match Test3) (D). Note different y-axes for sample and test period. The medians (indicated by the vertical lines) were tested against zero (no tuning). ***p < 0.001; *p < 0.05; n.s. not significant. (E-H) Probability density functions of cross-correlation coefficients (CC_{Tuning}) calculated for numerosity selective VIP neurons (N = 34) across Test periods. Same layout as in (A-D).

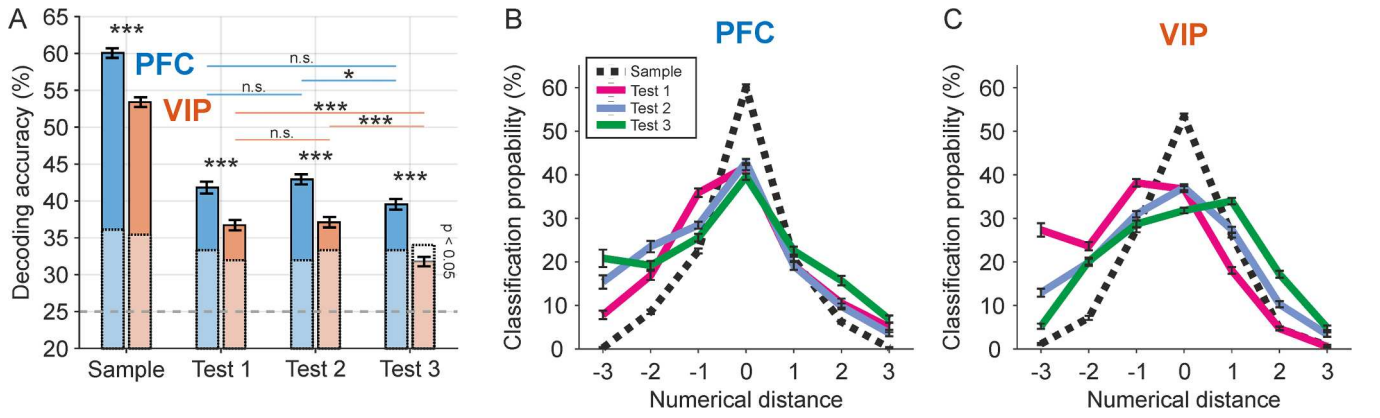


Fig. 11. Decoding numerosity with a SVM classifier from numerosity-selective neurons across match Test1 to Test3 periods. (A) Mean decoding accuracy for sample phase, and match Test1 to Test3 phases. The classifier was trained on sample activity and then tested on sample activity (odd versus event trials cross validation), and match activity in Test1, Test2 and Test3 phases. PFC and VIP cell counts were balanced by creating two pseudo populations of equal neuron numbers. In PFC, decoding accuracy remains stable and significantly better than in VIP throughout the Test periods. In VIP, numerosity decoding was no longer significant in the Test3 period. Layout as in Fig. 5A. (B) Normalized and averaged numerosity tuning curves derived from classifier performance on PFC neurons (N = 34) during the sample and match conditions of Test1, Test2, and Test3 phases. Classification probability for each resample is aligned to the preferred numerosity at center value 0 and plotted as a function of numerical distance. Error bars show the SEM across 100 resamples. Layout as in Fig. 5B. (C) Tuning functions from classifier performance on VIP neurons. Layout as in Fig. 5C.

for all test periods: $p < 0.001$, with alpha correction).

Finally, we derived tuning curves from the classifier performance, plotted them as a function of numerical distance aligned to the preferred numerosity, fitted Gauss functions to the tuning curves, and compared the sigma-values as a measure of tuning width (Fig. 11B,C). During the sample, Test1 and Test2 period, VIP showed significantly wider tuning functions compared to PFC (median sample PFC: $\sigma = 1.01$, median sample VIP: $\sigma = 1.27$, Wilcoxon rank sum test, $Z = -6.8171$, $p < 0.001$; median Test1 PFC: $\sigma = 1.86$, median Test1 VIP: $\sigma = 2.50$, Wilcoxon rank sum test, $Z = -5.5893$, $p < 0.001$; median Test2 PFC: $\sigma = 1.96$, median Test2 VIP: $\sigma = 2.26$, Wilcoxon rank sum test, $Z = -2.6987$, $p = 0.007$), with a tendency for sharper PFC tuning in the Test3 period (median Test3 PFC: $\sigma = 2.34$, median Test3 VIP: $\sigma = 2.58$, Wilcoxon rank sum test, $Z = -1.7702$, $p = 0.0767$)

3. Discussion

Previous studies have identified neural signatures for purely sensory recognition memory in a variety of high-level cortical areas in primates, such as the inferior temporal cortex (Baylis and Rolls, 1987; Riches et al., 1991; Eskandar et al., 1992; Miller et al., 1991, 1993; Li et al., 1993; Miller and Desimone, 1993, 1994; Vogels et al., 1995; Anderson et al., 2008; Woloszyn and Sheinberg; 2012; Meyer and Rust, 2018), the medial temporal lobe areas (Riches et al., 1991; Fahy et al., 1993), the posterior parietal cortex (Constantinidis and Steinmetz, 2001; Rawley and Constantinidis, 2010), and the lateral prefrontal cortex (Miller et al., 1996; Rainer and Miller, 2000; Freedman et al., 2003; Ng et al., 2014; Parto Dezfouli et al., 2021). In these regions, many neurons respond differently to test stimuli based on whether they match the sample; in other words, whether the monkey is memorizing the sample and attentively looking for it across time.

In the current study, we compared and contrasted the neuronal mechanisms of recognition memory involved in match versus nonmatch computation of abstract numerical quantities in monkeys. To perform the delayed match-to-numerosity task with variable frequencies of non-matching numerosities, the monkey had to retain a memory of the sample quantity and determine whether one of a variable number of test stimuli matches it. Within an ecological context, animals benefit from such a search image for numerosities, allowing them to identify cued numerical quantities — such as food, friends, or foes (Ishii and Shimada, 2010) — within a stream of numerical values (Nieder, 2020). Using analyses of single-neuron tuning and population decoding approaches, we identified aspects of recognition memory for quantities in the lateral PFC and the VIP within the intraparietal sulcus. The observed PFC effects are unlikely to be related to long-term memory, as the monkeys evaluated different numerosities on a trial-by-trial basis in the current DMS task. Instead, the sustained numerosity-selective activity in PFC likely reflects its role in working memory, consistent with evidence that PFC neurons maintain information over short delays to guide behavior in tasks requiring temporary numerical storage and manipulation. Below, we discuss the prominent differences between these two brain areas, which are known to form the primate core number network (Nieder and Miller, 2004; Nieder and Dehaene, 2009; Nieder, 2016).

In the PFC, we observed only mild modification of neuronal activity when the same preferred numerosity was repeated. Specifically, there was no degradation in tuning selectivity when the same sample numerosity was presented again as a matching Test stimulus. This indicates that PFC neurons maintained their numerosity selectivity between these trial phases. This held true even when the presentation of the target numerosity was interrupted and delayed by one or more non-matching numerosities. However, decoding accuracy for the population of selective PFC neurons dropped during the subsequent matching Test period, indicating some loss of selectivity with time. However, despite this initial drop, accuracy remained significantly and constantly above chance even when the target numerosity presentation was interrupted and postponed by non-matching numerosities. The observed reduction in neuronal activity when the sample stimulus is immediately repeated as the matching test stimulus is consistent with the phenomenon known as repetition suppression, which has been well documented for visual stimuli (Miller et al., 1991, 1993; Miller and Desimone, 1994; Steinmetz and Constantinidis, 1995; Constantinidis and Steinmetz, 2001). Our findings demonstrate that this effect also applies to abstract quantitative categories, such as the number of dots in a set.

Unlike the stable tuning to sought-after target numerosities, the tuning selectivity of PFC neurons specifically decreased when the same numerosities were presented as non-matches during the test periods. This decline was evident in both tuning selectivity and population decoding accuracy. The decoding analysis shows that only PFC neurons discriminate match from nonmatch numerosities, highlighting the PFC's sensitivity to behavioral relevance (Viswanathan and Nieder, 2015) and its role in distinguishing task-relevant quantities. When numerosities are being searched for as matches, PFC neurons respond with high selectivity and are finely tuned to them. However, tuning and decoding accuracy significantly decay when the same numerosities are not the ones

memorized and attended to by the monkey.

In contrast to PFC neurons, VIP neurons displayed three notable distinctions when monkeys were seeking a matching stimulus within a sequence of numerosities. First, VIP neurons exhibited an overall diminished responsiveness to numerosities across all trial periods, evident from the lower accuracy in population decoding analysis.

Second, the tuning selectivity of VIP neurons notably waned with repetitions of numerosities, irrespective of whether they were presented as matching or non-matching test stimuli. This indicates a lack of distinctions between match and non-match numerosities in VIP neurons. Consequently, the robust match-versus-nonmatch effect observed in PFC, as indicated by the match-versus-nonmatch index, was notably absent in VIP. VIP neurons seem to possess limited capabilities in discerning sought-after versus irrelevant numerosities.

Third, we observed differences in the mechanisms of how PFC and VIP neurons responded to repeated numerosities. In our study with numerical categories, significant match enhancement effects were found only in the PFC. Match enhancement is the effect of increased neuronal activity when a stimulus matches a previously encountered stimulus that has been stored in memory, an active mechanism that engages in favor of sought-after stimuli, ensuring that repetitive, irrelevant non-match stimuli are ignored (Engel and Wang, 2011). The presence of match enhancement effects exclusively in the PFC, alongside match suppression effects in both the VIP and PFC, supports the notion that the PFC plays a specialized role in enhancing the recognition of familiar numerosities. Match enhancement has been individually observed in the prefrontal (Miller et al., 1996; Freedman et al., 2003; Qi et al., 2012; Parto Dezfouli et al., 2021), temporal (Miller and Desimone, 1994), and parietal (Rawley and Constantinidis, 2010) cortices. Conversely, significant match suppression effects occurred in both the PFC and VIP, with similar magnitudes in these two brain areas. This reduction may help select PFC and VIP neurons filter out familiar numerosities, allowing them to focus on novel stimuli. Together, match enhancement and match suppression create a balance that enables PFC neurons, in particular, to efficiently process, store, and respond to numerical quantity by enhancing recognition of familiar numerosities while suppressing redundancy.

Overall, VIP activity lacks the capacity to discern and encode sought-after versus irrelevant numerosities and appears unable to track the numerosity the monkey is searching for. In contrast, PFC signals distinguish target versus irrelevant numerosities, serving as a signature for recognition memory. The observed stability of the tuning of PFC neurons during match conditions, in contrast to the overall less stable tuning of VIP neurons, highlights the specialized role of the PFC in maintaining precise numerical representations. The observed neuronal response latencies between these regions (see also Viswanathan and Nieder, 2013, 2015), and corroborating previous findings on earlier numerosity selectivity in VIP (Nieder and Miller, 2004) suggests that the VIP functions upstream of the PFC. This temporal sequencing underscores the hierarchical processing dynamics where VIP inputs precede and likely influence the activity within the PFC, thereby suggesting a feedforward mechanism in the neural circuitry. This aligns with established neuroanatomical pathways and functional connectivity patterns (Chafee and Goldman-Rakic, 2000; Crowe et al., 2013), reinforcing the role of VIP in preliminary numerical integration before the executive processing in the PFC. Overall, the data suggest that VIP neurons handle the initial encoding of numerosity in early cortical processing, routing numerical information to PFC neurons, which maintain the target number across temporal delays for recognition in delayed searches.

Conversely, the PFC is more involved in executive functioning with categories (Roy et al., 2010; Antzoulatos and Miller, 2011). This is also evident in the current study on number categories, in which monkeys had to indicate whether two sequentially presented numerosities were the same or different, and numerosity tuning for targets was stable across time whereas it deteriorated for non-matches. This correlates with a previous finding in a decision task in which PFC neurons initially represented numerical values with high precision before the neuronal codes became transformed during the course of the abstract decision process (Viswanathan et al., 2024). Together, these observations suggest that the PFC may function as a search image for numbers, enabling primates to swiftly and accurately identify specific numerical quantities within a sequence of numbers (Ishii and Shimada, 2010).

The neuronal patterns observed in the current study in PFC and VIP contrast with those from a previous numerosity distractor task (Jacob and Nieder, 2014; Jacob et al., 2018). In the classic distractor task, an irrelevant stimulus is passively viewed at a predefined trial period. However, in the current study, monkeys were presented with a variable count of sequential test displays, requiring active evaluation of each numerosity presentation to determine if it matched the number of dots. The absence of distractors in our study is a strength, as it isolates the core processes involved without the additional complexity of attention-related factors. This difference in task design may explain why PFC neuron tuning was affected by a passively viewed distractor (Jacob and Nieder, 2014; Jacob et al., 2016; Lin et al., 2023), whereas it remained stable in the current study where numerosity presentations were unpredictable. Conversely, VIP neuron tuning was unaffected by a passive distractor (Jacob and Nieder, 2014), but in the current task, VIP neurons appear unable to track the numerosity that the monkey is actively searching for. This suggests the sequential search task engages different cognitive processes and memory systems, emphasizing continuous comparison and evaluation of each display rather than suppressing an interfering stimulus.

The fronto-parietal network plays a crucial role in recognizing auditory numerical information. Both macaques and humans represent the number of tones in intraparietal and dorsal premotor areas, while tone repetition patterns are processed in ventral prefrontal cortex and basal ganglia (Wang et al., 2015). Macaques can detect first-order sequence violations in auditory cortex and second-order violations in a frontoparietal network, indicating a hierarchical predictive coding system similar to humans (Uhrig et al., 2014). These neural mechanisms are essential for recognition memory, enabling the encoding, maintenance, and retrieval of auditory information. The intraparietal and dorsal premotor areas contribute to the initial representation and maintenance of numerical and auditory information, while the ventral prefrontal cortex and basal ganglia are involved in processing and updating tone repetition patterns. The ability to detect sequence violations reflects the brain's capacity to predict and recognize auditory patterns, supporting the function of the recognition memory system.

Our study confirms that the IPS, particularly areas like LIP and MIP, play key roles in perceptual decision-making. In studies by Freedman and co-workers with monkeys performing a delayed match-to-category task involving visual motion direction, LIP neurons encoded category information with greater strength and faster latency than PFC neurons (Freedman and Assad, 2006; Swaminathan et al., 2012). While PFC primarily guided match/non-match decisions, LIP and MIP were more involved in stimulus evaluation and motor planning (Zhou et al., 2021). Pharmacological inactivation studies confirmed LIP's causal, task-independent role in memory-based visual categorical decisions, regardless of task structure or response modality (Zhou et al., 2023).

While these studies confirm the involvement of IPS regions in categorization and decision-making, they have not tested explicit recognition memory with multiple sequential choices, as examined in the current study. Additionally, the role of VIP, a key focus of the current work, has not been explored in motion direction categorization. Another crucial difference lies in the stimuli: Freedman and colleagues used arbitrary, learned motion direction categories, while we employed empirically ordered categories—the number of items in a set—aligned along a "number line." Despite these differences, both approaches highlight the importance of the IPS within the fronto-parietal categorization network.

Research on neuronal responses in the monkey ventral intraparietal

area (VIP) suggests a complex relationship with motor behavior. VIP neurons show sensitivity to visual motion and heading direction, with responses modulated by attention and eye movements (Cook and Maunsell, 2002; Zhang and Britten, 2010, 2011). These findings collectively suggest that VIP neuronal responses are influenced by motor behavior and task context.

The macaque ventral intraparietal area (VIP) is involved in various sensorimotor and cognitive functions, including motion processing, multisensory integration, head peripersonal space processing, defensive behavior, and numerosity coding (Foster et al., 2022). In humans, VIP has evolved into a "VIP complex" comprising three distinct parietal areas with specialized functions related to self-representation and spatial processing. A proposed unifying functional principle is "prediction in space and time," linking VIP to state estimation, potentially including numerical recognition memory, as a key parietal sensorimotor function (Foster et al., 2022).

In our proposed hierarchical model, VIP neurons are responsible for the initial encoding of numerosity, while PFC neurons maintain target numbers across temporal delays for recognition. To further enrich this framework, we propose incorporating the hippocampus and MTL structures as critical components in binding numerical information with contextual details, thereby facilitating the formation and retrieval of robust numerical memories. The hippocampus supports recognition memory by retrieving studied details, while the MTL cortex provides familiarity signals (Norman and O'Reilly, 2003). Both the MTL and PFC are crucial for visual short-term memory of numerical information (Kutter et al., 2018), with their interactions underlying working memory mechanisms (Wu and Buckley, 2022), including arithmetic memory (Menon, 2016). This integration is consistent with established neuroanatomical pathways and functional connectivity patterns, underscoring the collaborative roles of these regions in numerosity processing and recognition memory.

Besides the fronto-parietal number network, medial temporal lobe (MTL) areas, including the hippocampus, are known to support sensory recognition memory (Riches et al., 1991; Fahy et al., 1993; Brown and Aggleton, 2001). In humans, MTL neurons also encode numerical values (Kutter et al., 2018, 2023, 2024) and working memory-related arithmetic rules (Kutter et al., 2022). Although traditionally linked to long-term memory, hippocampal neurons can show sustained, feature-selective delay activity predicting successful working memory retrieval (Kornblith et al., 2017; Kamiński et al., 2017). While not directly assessed here, the hippocampus and MTL likely contribute to a broader network, working alongside the PFC and VIP to support numerical cognition and recognition memory.

While numerous computational models of recognition memory have been proposed (Sohal and Hasselmo, 2000; Norman and O'Reilly, 2003; Bogacz and Brown, 2003; Cortes et al., 2010; Tyulmankov et al., 2022), most of these models treat familiarity discrimination as a binary problem for unrelated stimuli — determining whether an arbitrary stimulus is novel or familiar. However, the situation is more complex for numerical quantities, which are represented in an orderly fashion on a mental number line. The bell-shaped tuning of numerosity-selective neurons and the resulting neuronal distance effect reflect the inherent relatedness between numerical values that are part of recognition memory for quantities. Neuronal network models have successfully simulated the emergence of numerosity tuning (Dehaene and Changeux, 1993; Verguts and Fias, 2004; Nasr et al., 2019; Nasr and Nieder, 2021; Mistry et al., 2023) and the counting of stimulus occurrence frequencies (Dasgupta et al., 2022). However, the neural circuit responsible for storing numerosity recognition memory has yet to be modeled.

The current physiological data are essential for advancing models of the numerosity system, as they help us move beyond the mere extraction of numerical quantity to understanding how the brain uses numerical information for goal-directed tasks, such as recognizing target numbers after a delay. Future studies could explore how task difficulty and training influence the neural correlates of numerosity recognition memory. Additionally, investigating the role of other brain areas, like the MTL regions or anterior inferotemporal cortex, would be a valuable direction for further research.

4. Methods

4.1. Experimental animals and surgery

Two male adult rhesus monkeys (*Macaca mulatta*) were implanted with two recording chambers each, centered over the principal sulcus in the dorsolateral prefrontal cortex (PFC) and the intraparietal sulcus (IPS) in the posterior parietal cortex (monkey 1 and monkey 2, both left hemisphere). Chamber implantation was guided by anatomical magnetic resonance imaging (MRI) and stereotaxic measurements. All surgeries were performed under sterile conditions while the monkeys were under general anesthesia. The monkeys received postoperative antibiotics and analgesics. All procedures were performed in accordance with the guidelines for animal experimentation approved by local authorities, the Regierungspräsidium Tübingen, Germany.

4.2. Experimental apparatus and stimuli

Monkeys were seated inside primate chairs in a chamber 57 cm away from a 15" flat screen monitor (with a resolution of 1024 by 768 pixels and a refresh rate of 60 Hz). The NIMH Cortex software (Bethesda, MD) was used to present the stimuli and behavioral data acquisition. The monkeys' eye gaze was tracked by an infrared eye tracking system (ISCAN, Cambridge, MA).

We used random dot arrays as numerosity stimuli ranging between 1 and 5 dots (diameter range from 0.5 ° to 0.9 ° of visual angle). Each unique dot pattern was presented on a gray background circle subtending 5.4 ° of visual angle and was used as sample and test stimulus. A custom-written MATLAB (Version R2020b, MathWorks Inc., Natick, MA) program generated the stimuli anew before each recording session to prevent the monkeys from memorizing stimulus sequences. Sample and test stimulus of the same countable numerosity were never identical within a trial and session. In order to control for low-level, non-numerical visual features in half of the trials, dot diameters were selected randomly (standard protocol). In the other half of the trials, dot density and total occupied area were equated across stimuli (control protocol). Trials of the standard and control protocol were pseudo-randomly and unpredictably presented.

4.3. Behavioral protocol

We trained two macaque monkeys on a modified version of the delayed match-to-numerosity task in which the monkeys had to match test numerosities to sample numerosities (Fig. 1A). Nonmatching numerosities that were presented between sample and matching test stimulus had to be ignored. Up to three sequential numerical nonmatches were presented, i.e. the task consisted of up to three test phases and in each test phase the right or wrong numerosity could occur. A trial was initiated by the monkeys when they grabbed a bar and maintained eye fixation within a window of 3.5° of visual angle, centered around a central fixation point, i.e., ±1.75° from the center. Trials were immediately aborted (blue screen appears) and excluded from further analysis if the animals broke fixation. At first, a gray circular background was shown for 500 ms (pure fixation period). Subsequently, the sample stimulus containing 1–4 dots appeared for 600 ms. This sample stimulus had to be memorized over a delay of 1000 ms (again gray background circle) and had to compared in a following sequence of test numerosities. Each test stimulus was presented for 800 ms with a 100 ms delay between each other. Since single trials with three test displays lasted nearly 5 s, during which the monkeys were required to maintain eye fixation, extending the delay intervals would have further prolonged the trials and led to a significant increase in fixation breaks, making it impractical

to incorporate longer delays.

For the case that the numerosity in test stimulus was same as in the sample (match situation, 1 – 4), the monkeys had to respond with a bar release. If the numerosities were different (nonmatch situation, 1 – 5), the animals continued to hold the bar. A nonmatch numerosity was never shown twice per trial. Sample numerosities and trial types (i.e. if no nonmatch, one nonmatch, two nonmatch or three nonmatch numerosities were presented) were balanced across conditions and displayed with equal probability (25 % each, pseudo-randomly intermixed). Thus, the monkeys could not predict when or whether the match stimulus would appear. Correct decision per trial was rewarded with a drop of water. In 25 % of the trials, when presenting three nonmatch numerosities in sequence, the monkeys were rewarded when holding the bar throughout the whole trial until display-offset. Chance level of 25 % was achieved by averaging the probabilities for each trial type (for details see Figure S1). Since the number of trials a monkey can perform during an experimental session is limited (on average monkey 1 completed 483 trials and monkey 2 534 trials), not all possible combinations of sample (1 – 4) and test numerosities (1 – 5) could be presented. Therefore, a predefined collection of 64 unique sample-test-combinations was shown.

4.4. Neurophysiological recordings

In each session, arrays of up to eight glass-coated 1 MΩ tungsten microelectrodes (Alpha Omega LTD, Israel) were inserted in each recording chamber using custom-made screw microdrives and a grid (Crist Instruments, USA) with 1 mm spacing. Stable and well isolated neurons were randomly selected; no attempt was made to preselect neurons according to response properties. To access the IPS, electrodes were passed along the course of it to a depth of 9–10 mm (Fig. 2A). A MAP Plexon system was used for signal acquisition, amplification, filtering and digitalization. Waveform separation was performed offline (Plexon Systems, USA).

4.5. Data analysis

All analyses were conducted in MATLAB (version R2021a, MathWorks Inc.) using custom-written scripts. All values in the main text and figures refer to the mean ± standard error of the mean (SEM), unless stated otherwise. SEM was calculated as the standard deviation divided by the square root of the number of samples.

Unless specified otherwise, neurons were included in the below listed analyses if the following criteria were met: First, their average firing rate across trials was at least 0.5 spikes/s; second, they were recorded for at least 18 correct trials per numerosity during the sample period (4 sample numerosities); and third, they were recorded for at least 3 correct trials per numerosity in each of the three test periods where a match numerosity occurred (match Test1 to Test3), as well as in the first test period when a nonmatch numerosity was presented (nonmatch Test1). To have the same numerical range between match and nonmatch situation (1 – 4 match vs. 1 – 4 nonmatch), we excluded all trials that contained nonmatch numerosity 5 in Test1 period. A total of 369 neurons in PFC and 254 neurons in VIP fulfilled these criteria.

4.6. Behavioral data analysis

For each recording sessions behavioral tuning functions were computed to describe the percentage of trials for which a test stimulus was judged as being equal in number to the sample. Mean tuning functions were computed as the average over all individual functions per session. Furthermore, the averaged performance for the four trial types (i.e. no nonmatch, one nonmatch, two nonmatch or three nonmatch test stimuli) according to the four sample numerosities was calculated to see an influence on the behavior. For each correct trial, in which the monkey had to release the bar (75 % of the trials), the reaction time was measured after match stimulus onset.

4.7. Visual response latency analysis

The visual response latency analysis was accomplished for all neurons with an average firing rate of at least 0.5 spike/s across trials and at least 5 correct trials per sample numerosity and stimulus protocol (4 sample numerosities x 2 stimulus protocols). For each cell the firing rate was averaged over all correct trials in 20 ms windows in 1 ms steps. Mean and standard deviation (SD) of baseline activity was measured for the last 200 ms of the presample period plus the first 40 ms of the sample period. The latency analysis was restricted to 40 ms till 300 ms after sample onset and was defined as the first time bin within this window in which the activity was higher or lower than 2 SDs compared to the baseline activity for more than 20 consecutive windows. Latencies estimated by visual inspection coincided with latencies determined in this way in most of the cases. We compared the latency distributions between VIP and PFC with a Wilcoxon rank sum test (two-sided). Furthermore, we calculated the 25th percentile of the respective latency distribution to take this value as the 'main' latency for all subsequent analyses. VIP neurons were analyzed from now on with a latency of 86 ms and PFC neurons with a latency of 100 ms after sample or test onset.

4.8. Neuronal selectivity and tuning analysis

Neuronal activity was analyzed separately for the sample period, for the first test period in which a nonmatch was presented and the three subsequent test periods showing the match situation. Since the monkeys were required to respond with a bar release when they detected a match numerosity, we only used correct trials in which the animals responded after 400 ms to ensure that the endpoint of the analysis window was not contaminated by motor activity. This was the case for 84.4 % of all possible match trials in monkey 1 and for 89.1 % in monkey 2. Trials with a shorter reaction time were excluded from further analysis. In the nonmatch situation such limitation was not necessary, since the monkeys had to withhold the bar throughout the nonmatch presentation. Thus, the time window for all analyses started with the 'main' latency of the respective brain area (see above) and ended 400 ms after sample or test onset.

To determine numerosity-selectivity during the sample period, a two-factorial analysis of variance (ANOVA) was performed with the factors numerosity (sample numerosity 1 – 4) and protocol (standard and control). Neurons with a significant main effect for the factor numerosity ($p < 0.01$), but no main effect for the factor protocol or interaction, were classified as exclusively numerosity-selective. This strict significance threshold of $p < 0.01$ was applied only for assessing firing rate differences using the ANOVA to avoid false positive numerosity selective neurons; for all other statistics, the conventional significance threshold of $p < 0.05$ was used. For numerosity-selective neurons, we created neuronal tuning functions for the sample, nonmatch Test1 period and match Test1 to Test3 period within the above described time window by averaging the firing rates across trials. The numerosity eliciting the highest discharge rate was defined as "most-preferred numerosity" of a given cell, whereas the numerosity provoking the lowest firing rate was referred as "least-preferred numerosity". Using this approach, we were also able to compute average and normalized neuronal tuning functions by averaging and normalizing (between 0 % and 100 %) all individual tuning functions with the same preferred sample numerosity.

4.9. Population-tuning analysis

To derive averaged numerosity-filter functions for the sample, for the nonmatch Test1 period and for the match Test1 to Test3 period, the individual tuning functions of each neuron were normalized by the ac-

tivity between its most-preferred and least-preferred numerosity determined during the sample period. Thus, the activity at the preferred numerosity for the sample period is set to 1. Furthermore, this maximum activity was used as the reference to compute the averaged numerosity-filter functions for the other test periods. In other words, the initially defined most-preferred numerosity remains the same across all test periods, even though the preferred numerosity may not necessarily correspond to the maximum activity response. We pooled the resulting normalized tuning curves across such neurons which were defined in the sample period as exclusively numerosity-selective by a two-way ANOVA ($p < 0.01$). The time windows to determine the neuronal tuning functions had equal length in sample and the respective test period ('main' visual response latency till 400 ms after sample or test onset). We compared the peak values of each averaged numerosity-filter function by using a permutation test (normalized firing rates, 10.000 permutations with replacement). Furthermore, we quantified tuning changes by fitting each normalized individual tuning function by a symmetric Gaussian distribution, described by the following formula:

$$f(x) = amp * e^{-\left(\frac{x-xc}{\sigma}\right)^2} + y0$$

To acquire the best fit to the data, σ (standard deviation of the Gaussian distribution), and *amp* (amplitude) were adjustable during the fitting procedure. The peak function's yaxis offset $y0$ was set to zero, and the center of the fitted distributions xc was fixed at the preferred numerosity that was determined during the sample period. The sample period served as the reference for determining each neuron's preferred numerosity, as there were no task demands during this time to influence tuning. Tuning curves from subsequent periods are aligned to this reference, and any deviation from the preferred numerosity in later test periods indicates distorted tuning. We used a Wilcoxon signed rank test (with Bonferroni correction, two-sided) to compare the Gaussian widths (σ) determined for the match and nonmatch situation, as well as for the match situations among themselves.

4.10. Index for tuning strength variations

To quantify tuning strength variations between the match and nonmatch test situation, we calculated for each cell an index for the most preferred numerosity determined during the sample period. The index can be calculated by subtracting the mean neuronal response to the match presentation (FR_{match}) from those to the nonmatch presentation (FR_{nonmatch}) and dividing the absolute value of the difference by the sum of the two means. It can be described by following formula:

$$Index = \frac{(FR_{match} - FR_{\text{nonmatch}})}{(FR_{\text{match}} + FR_{\text{nonmatch}})}$$

Neurons that responded stronger to a numerosity when displayed as a match compared to when displayed as sample showed a match-enhancement effect. Neurons that responded stronger to a numerosity when displayed as a sample compared to when displayed as nonmatch showed a match suppression effect. For a better visualization, we calculated for cells that show a match suppression effect the absolute value, whereby the index can range only between 0 and 1. Values close to 0 indicate equal responses to the two respective test situations.

We used this calculation for two approaches: First, we pooled the absolute indices to see general tuning strength variations between the match and nonmatch situation of the Test1 period (match Test1 vs. nonmatch Test1), but also for comparing the match situation of the Test2 and Test3 period with the nonmatch situation of the Test1 period (match Test2 vs. nonmatch Test1, match Test3 vs. nonmatch Test1). Second, we went more into detail and calculated the indices separately for the match enhancement and match suppression effect of the Test1 and Test2 period (match vs. nonmatch). To see which tuning strength variation (match enhancement or match suppression) can explained by chance, we calculated a shuffled distribution by dissolving the association between the respective test situations (i.e. match or nonmatch) and the trial-by-trial firing rates of each neuron. We repeated this process 10,000 times, calculated the mean of these single values and compared the median of the new distribution with the median of the real data by a Wilcoxon signed rank test (two-sided). Furthermore, we tested with a one-sided Wilcoxon rank sum test if the median PFC index is larger than the median VIP index. For this analysis we considered just neurons, which were in the sample period determined as exclusively numerosity-selective by a two-way ANOVA ($p < 0.01$). Firing rates were derived from time windows with equal length in sample and the respective test period ('main' visual response latency till 400 ms after sample or test onset).

4.11. Tuning curve cross-correlation analysis

Cross-correlations were performed between neuronal tuning functions (Diester and Nieder, 2007) calculated during the sample period and the subsequent three match periods. This kind of analysis provides a measure to quantify the extent to which single neurons changes their initial coding (i.e. representing the sample numerosities) during the presentation of test numerosities. Furthermore, we calculated the cross-correlation for the sample period for itself by comparing the tuning functions derived from odd trial numbers with that evaluated for even trial numbers. We considered just neurons, which were in the sample period determined as exclusively numerosity-selective by a two-way ANOVA ($p < 0.01$). Tuning curves were derived from time windows with equal length in sample and the respective test period ('main' visual response latency till 400 ms after sample or test onset). The cross-correlations were calculated by following formula:

$$\text{CCTuning} = \frac{\sum_{n=1}^{4}\left(FR_{\text{sample}(n)} - \overline{FR}_{\text{sample}}\right) * \left(FR_{\text{test}(n)} - \overline{FR}_{\text{test}}\right)}{\sqrt{\sum_{n=1}^{4}\left(FR_{\text{sample}(n)} - \overline{FR}_{\text{sample}}\right)^2} * \sqrt{\sum_{n=1}^{4}\left(FR_{\text{test}(n)} - FR\text{t}_{\text{es}}FR\text{t}\right)^2}},$$

where $FR_{\text{sample}}(n)$ is the average firing rate for numerosity $n \in \{1,2,3,4\}$ during the sample period, $\overline{FR}_{\text{sample}}$ the average firing rate across all numerosities during the sample period, $FR_{\text{test}}(n)$ the average firing rate for numerosity $n \in \{1,2,3,4\}$ during the test period, and $\overline{FR}_{\text{test}}$ the average firing rate across all numerosities during the test period. Thus, the higher the cross-correlation coefficients, the more stable is the initial encoding across the respective test period. We performed this analysis for VIP and PFC neurons separately, resulting in four different cross-correlation coefficient distributions per brain area (sample and three match periods). Each distribution was tested against zero by bootstrapping the medians 10,000 times with replacement. The significance threshold was set to $p < 0.05$, i.e. the 5th and 95th percentile. For a better comparison between the brain areas, the probability density function was plotted as a histogram.

4.12. SVM classifier

We trained a linear multiclass support vector machine (SVM) classifier (one-versus-one classification, Chang and Lin, 2011) on trial firing rates during the sample period and tested it for the discrimination of numerosities on spiking activity derived from the sample period itself, from the nonmatch Test1 period and from the subsequent three match periods. This resulted in five analyses per brain area. We only used neurons which were determined as exclusively-numerosity selective during the sample period. A six-fold cross-validation was performed, resulting in 15 trials for training and 3 trials for testing per numerosity. Within each cross-validation loop a new training and test subset was drawn. However, through the limitation of trials, it was possible that for the test subset derived from the test periods the same trials were drawn multiple times. Trial firing rates were z-scored on the parameters from the training subset. We repeated training and testing 100 times, selecting with each iteration a new set of trials randomly. The tested SVM

classifier provides the predicted class labels for the trials in the test subset. This allows us to compute the main decoding success as well as to construct in combination with the true class labels of the test subset confusion matrices. Both possess an accuracy measure in percent. From each confusion matrix we calculated the classification probability as a function of distance from the real numerosity derived from the diagonals (averaged performance parallel to the mean diagonal). The resulting classification probability curve was fitted by a symmetric Gaussian distribution to derive sigma-values as a measure of tuning width for each iteration. It can be described by the following formula:

$$f(x) = amp * e - \left(\frac{x - xc}{\sigma}\right)^2 + y0$$

The standard deviation of the Gaussian distribution (σ) was adjustable during the fitting procedure. The peak function's yaxis offset $y0$ was set to zero, the amplitude (amp) corresponded to the classification probability of the mean diagonal and the center (xc) of the fitted distribution was fixed at zero. We used a Wilcoxon rank sum test (two-sided) to compare the Gaussian widths (σ) determined for each situation.

We evaluated the decoding performance of the SVM classifier in balanced VIP and PFC pseudo-populations by drawing with each iteration a new set of neurons with equal size (decisive factor was the lowest neuron number of the respective brain area). In order to create a chance level performance distribution, we randomly shuffled the association between firing rates and trial class labels (numerosity) with each iteration. The significance threshold was set to $p < 0.05$ (one-sided), i.e. the actual decoding success value was required to be larger than 95 % of values in the shuffled distribution. We tested decoding success distributions between VIP and PFC as well as between sample, match and nonmatch with a permutation test (with alpha correction, 10.000 permutations with replacement) for each stimulus period.

4.13. PEV analysis

To quantify the information about the match (same as sample) or nonmatching (other than sample) numerosity that was carried by a neuron over the time course, we performed the percent explained variance (ω^2 PEV) anaylsis. As other studies already showed (e.g. Buschman et al., 2011; Jacob and Nieder, 2014), the PEV analysis is suitable to measeure the amount of information carried by neuronal populations. In our case, PEV reflects how much of the variance in a neuron's firing rate can be explained by the match and nonmatch numerosity. For this, we considered all neurons that were recorded with at least 18 correct trials per sample numerosity and 3 correct trials per test numerosity in each test phase (match/nonmatch). 369 PFC neurons and 254 VIP neurons fullfilled this criterion. We used a one-way categorical sliding-window ANOVA (200 ms window with 20 ms step size) over the relevant time window of the trial (baseline onset until 400 ms after test onset) to yield the sum-of-squares. From these, we calculated ω^2 according to the following formula:

$$\omega^2 = \frac{SS_{Groups} - df * MSE}{SS_{Total} + MSE} * 100,$$

where SS_{Groups} denotes the sum-of-squares between groups (numerosities), SS_{Total} the total sum-of-squares, df the degrees of freedom, and MSE the mean squared error. The number of trials in each group was balanced to 3 trials. For this, a random subset of trials was drawn and the statistic was calculated. This process was repeated 25 times, and the overall statistic was taken to be the mean of the stratified values. The resulting values were averaged over units to extract the population PEV as a function of time for match and nonmatch numerosities. Furthermore, we calculated the mean PEV during the Test1 period for match and nonmatch condition for a better comparison. For this, we used firing rates which were derived from time windows between the 'main' visual response latency for the respective brain area and 400 ms after test onset. Then, the same calculation process as described above followed. In order to create a baseline PEV, we randomly shuffled the association between firing rates and numerosity. This process was repeated 1000 times, and the shuffled PEV was taken to be the mean of the individual values. The significance threshold was set to $p < 0.01$ (one-sided), i.e. the PEV was required to be larger than 99 % of values in the shuffled distribution. Bin-wise Wilcoxon signed rank tests (evaluated at $p < 0.05$) were used to compare match and nonmatch numerosity PEV within the population of PFC and VIP neurons.

Author contributions

Conceptualization and Methodology: TM, AN; Investigation, Data curation, Analysis, Software and Visualization: TM, JG; Writing – Original draft: TM, AN; Writing - Review/Editing: TM, JG, AN; Funding acquisition and Supervision: AN.

CRediT authorship contribution statement

Julia Grüb: Writing – review & editing, Investigation, Formal analysis. **Tobias Machts:** Writing – review & editing, Writing – original draft, Software, Investigation, Formal analysis, Conceptualization. **Andreas Nieder:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by DFG grant NI 618/10–1 to A.N.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.pneurobio.2025.102794.

Data availability

Data will be made available on request.

References

Anderson, B., Mruczek, R.E., Kawasaki, K., Sheinberg, D., 2008. Effects of familiarity on neural activity in monkey inferior temporal lobe. Cereb. Cortex 18 (11), 2540–2552.
Antzoulatos, E.G., Miller, E.K., 2011. Differences between neural activity in prefrontal cortex and striatum during learning of novel abstract categories. Neuron 71 (2), 243–249.
Baylis, G.C., Rolls, E.T., 1987. Responses of neurons in the inferior temporal cortex in short term and serial recognition memory tasks. Exp. Brain Res. 65, 614–622.
Bogacz, R., Brown, M.W., 2003. Comparison of computational models of familiarity discrimination in the perirhinal cortex. Hippocampus 13 (4), 494–524.
Brown, M.W., Aggleton, J.P., 2001. Recognition memory: what are the roles of the perirhinal cortex and hippocampus? Nat. Rev. Neurosci. 2, 51–61.
Buckley, P.B., Gillman, C.B., 1974. Comparisons of digits and dot patterns. J. Exp. Psychol. 103 (6), 1131–1136.
Buschman, T.J., Siegel, M., Roy, J.E., Miller, E.K., 2011. Neural substrates of cognitive capacity limitations. Proc. Natl. Acad. Sci. 108, 11252–11255.
Chafee, M.V., Goldman-Rakic, P.S., 2000. Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. J. Neurophysiol. 83 (3), 1550–1566.
Chang, C.-C., Lin, C.-J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 27.
Constantinidis, C., Steinmetz, M.A., 2001. Neuronal responses in area 7a to multiple stimulus displays: II. Responses are suppressed at the cued location. Cereb. Cortex 11, 592–597.

Cook, E.P., Maunsell, J.H., 2002. Attentional modulation of behavioral performance and neuronal responses in middle temporal and ventral intraparietal areas of macaque monkey. J. Neurosci. 22 (5), 1994–2004.
Cortes, J.M., Greve, A., Barrett, A.B., van Rossum, M.C., 2010. Dynamics and robustness of familiarity memory. Neural Comput. 22 (2), 448–466.
Crowe, D.A., Goodwin, S.J., Blackman, R.K., Sakellaridi, S., Sponheim, S.R., MacDonald 3rd, A.W., Chafee, M.V., 2013. Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. Nat. Neurosci. 16 (10), 1484–1491.
Dasgupta, S., Hattori, D., Navlakha, S., 2022. A neural theory for counting memories. Nat. Commun. 13, 5961.
Dehaene, S., Bossini, S., Giraux, P., 1993. The mental representation of parity and number magnitude. J. Exp. Psychol. Gen. 122, 371–396.
Dehaene, S., Changeux, J.P., 1993. Development of elementary numerical abilities: a neuronal model. J. Cogn. Neurosci. 5 (4), 390–407.
Diester, I., Nieder, A., 2007. Semantic associations between signs and numerical categories in the prefrontal cortex. PLoS Biol. 5 (11), e294.
Engel, T.A., Wang, X.J., 2011. Same or different? A neural circuit mechanism of similarity-based pattern match decision making. J. Neurosci. Official J. Soc. Neurosci. 31 (19), 6982–6996.
Eskandar, E.N., Richmond, B.J., Optican, L.M., 1992. Role of inferior temporal neurons in visual memory. I. Temporal encoding of information about visual images, recalled images, and behavioral context. J. Neurophysiol. 68, 1277–1295.
Fahy, F.L., Riches, I.P., Brown, M.W., 1993. Neuronal activity related to visual recognition memory: long-term memory and the encoding of recency and familiarity information in the primate anterior and medial inferior temporal and rhinal cortex. Exp. Brain Res. 96 (3), 457–472.
Foster, C., Sheng, W.A., Heed, T., Ben Hamed, S., 2022. The macaque ventral intraparietal area has expanded into three homologue human parietal areas. Prog. Neurobiol. 209, 102185.
Freedman, D.J., Assad, J.A., 2006. Experience-dependent representation of visual categories in parietal cortex. Nature 443, 85–88.
Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2003. A comparison of primate prefrontal and inferior temporal cortices during visual categorization. J. Neurosci. 23, 5235–5246.
Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K., 2006. Experience-dependent sharpening of visual shape selectivity in inferior temporal cortex. Cereb. Cortex 16 (11), 1631–1644.
Galton, F., 1880. Visualised numerals. Nature 21, 252–256.
Ishii, Y., Shimada, M., 2010. The effect of learning and search images on predator–prey interactions. Popul Ecol. 52, 27–35.
Jacob, S.N., Hähnke, D., Nieder, A., 2018. Structuring of abstract working memory content by fronto-parietal synchrony in primate cortex. Neuron 99 (3), 588–597 e5.
Jacob, S.N., Nieder, A., 2014. Complementary roles for primate frontal and parietal cortex in guarding working memory from distractor stimuli. Neuron 83 (1), 226–237.
Jacob, S.N., Stalter, M., Nieder, A., 2016. Cell-type-specific modulation of targets and distractors by dopamine D1 receptors in primate prefrontal cortex. Nat. Commun. 7, 13218.
Kamiński, J., Sullivan, S., Chung, J.M., Ross, I.B., Mamelak, A.N., Rutishauser, U., 2017. Persistently active neurons in human medial frontal and medial temporal lobe support working memory. Nat. Neurosci. 20, 590–601.
Kornblith, S., Quian Quiroga, R., Koch, C., Fried, I., Mormann, F., 2017. Persistent single-neuron activity during working memory in the human medial temporal lobe. Curr. Biol. 27, 1026–1032.
Kutter, E.F., Boström, J., Elger, C.E., Mormann, F., Nieder, A., 2018. Single neurons in the human brain encode numbers. Neuron 100 (3), 753–761 e4.
Kutter, E.F., Boström, J., Elger, C.E., Nieder, A., Mormann, F., 2022. Neuronal codes for arithmetic rule processing in the human brain. Curr. Biol. CB 32 (6), 1275–1284 e4.
Kutter, E.F., Dehnen, G., Borger, V., Surges, R., Mormann, F., Nieder, A., 2023. Distinct neuronal representation of small and large numbers in the human medial temporal lobe. Nat. Hum. Behav. 7 (11), 1998–2007.
Kutter, E.F., Dehnen, G., Borger, V., Surges, R., Nieder, A., Mormann, F., 2024. Single-neuron representation of nonsymbolic and symbolic number zero in the human medial temporal lobe. Curr. Biol. CB 34 (20), 4794–4802 e3.
Li, L., Miller, E.K., Desimone, R., 1993. The representation of stimulus familiarity in anterior inferior temporal cortex. J. Neurophysiol. 69 (6), 1918–1929.
Lim, S., McKee, J.L., Woloszyn, L., Amit, Y., Freedman, D.J., Sheinberg, D.L., Brunel, N., 2015. Inferring learning rules from distributions of firing rates in cortical neurons. Nat. Neurosci. 18 (12), 1804–1810.
Lin, X.X., Nieder, A., & Jacob, S.N. (2023). The neuronal implementation of representational geometry in primate prefrontal cortex. Science advances, 9(50), eadh8685.
Menon, V., 2016. Memory and cognitive control circuits in mathematical cognition and learning. Prog. Brain Res. 227, 159–186.
Merten, K., Nieder, A., 2009. Compressed scaling of abstract numerosity representations in adult humans and monkeys. J. Cogn. Neurosci. 21 (2), 333–346.
Meyer, T., Rust, N.C., 2018. Single-exposure visual memory judgments are reflected in inferotemporal cortex. eLife 7, e32259.
Miller, E.K., Desimone, R., 1994. Parallel neuronal mechanisms for short-term memory. Science 263, 520–522.
Miller, E.K., Erickson, C.A., Desimone, R., 1996. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J. Neurosci. 16, 5154–5167.
Miller, E.K., Li, L., Desimone, R., 1991. A neural mechanism for working and recognition memory in inferior temporal cortex. Science 254, 1377–1379.
Miller, E.K., Li, L., Desimone, R., 1993. Activity of neurons in anterior inferior temporal cortex during a short-term memory task. J. Neurosci. 13, 1460–1478.
Mistry, P.K., Strock, A., Liu, R., Young, G., Menon, V., 2023. Learning-induced reorganization of number neurons and emergence of numerical representations in a biologically inspired neural network. Nat. Commun. 14 (1), 3843.
Moyer, R.S., Landauer, T.K., 1967. Time required for judgements of numerical inequality. Nature 215 (2), 1519–1520.
Nasr, K., Nieder, A., 2021. Spontaneous representation of numerosity zero in a deep neural network for visual object recognition. iScience 24 (11), 103301.
Nasr, K., Viswanathan, P., Nieder, A., 2019. Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. Sci. Adv. 5 (5) eaav7903.
Ng, C.W., Plakke, B., Poremba, A., 2014. Neural correlates of auditory recognition memory in the primate dorsal temporal pole. J. Neurophysiol. 111 (3), 455–469.
Nieder, A., 2016. The neuronal code for number. Nat. Rev. Neurosci. 17 (6), 366–382.
Nieder, A., 2020. The adaptive value of numerical competence. Trends Ecol. Evol. 35 (7), 605–617.
Nieder, A., 2023. Convergent circuit computation for categorization in the brains of primates and songbirds. Cold Spring Harb. Perspect. Biol. 15 (12), a041526.
Nieder A. (2024). The Calculating Brain. Physiological reviews, 10.1152/physrev.00014.2024. Advance online publication. https://doi.org/10.1152/physrev.00014.2024.
Nieder, A., Dehaene, S., 2009. Representation of number in the brain. Annu. Rev. Neurosci. 32, 185–208.
Nieder, A., Diester, I., Tudusciuc, O., 2006. Temporal and spatial enumeration processes in the primate parietal cortex. Science 313 (5792), 1431–1435.
Nieder, A., Freedman, D.J., Miller, E.K., 2002. Representation of the quantity of visual items in the primate prefrontal cortex. Science 297 (5587), 1708–1711.
Nieder, A., Merten, K., 2007. A labeled-line code for small and large numerosities in the monkey prefrontal cortex. J. Neurosci. 27 (22), 5986–5993.
Nieder, A., Miller, E.K., 2003. Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. Neuron 37 (1), 149–157.
Nieder, A., Miller, E.K., 2004. A parieto-frontal network for visual numerical information in the monkey. Proc. Natl. Acad. Sci. 101 (19), 7457–7462.
Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol. Rev. 110 (4), 611–646.
Norman, K.A., O'Reilly, R.C., 2003. Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol. Rev. 110 (4), 611–646.
Okuyama, S., Kuki, T., Mushiake, H., 2015. Representation of the numerosity 'zero' in the parietal cortex of the monkey. Sci. Rep. 5, 10059.
Parto Dezfouli, M., Zarei, M., Constantinidis, C., Daliri, M.R., 2021. Task-specific modulation of PFC activity for matching-rule governed decision-making. Brain Struct. Funct. 226 (2), 443–455.
Qi, X.-L., Meyer, T., Stanford, T.R., Constantinidis, C., 2012. Neural correlates of a decision variable before learning to perform a match/non-match task. J. Neurosci. 32, 6161–6169.
Rainer, G., Miller, E.K., 2000. Effects of visual experience on the representation of objects in the prefrontal cortex. Neuron 27 (1), 179–189.
Rawley, J.B., Constantinidis, C., 2010. Effects of task and coordinate frame of attention in area 7a of the primate posterior parietal cortex. J. Vis. 10 (1), 1–16.
Riches, I.P., Wilson, F.A., Brown, M.W., 1991. The effects of visual stimulation and memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. J. Neurosci. 11, 1763–1779.
Riches, I.P., Wilson, F.A., Brown, M.W., 1991. The effects of visual stimulation and memory on neurons of the hippocampal formation and the neighboring parahippocampal gyrus and inferior temporal cortex of the primate. J. Neurosci. Official J. Soc. Neurosci. 11 (6), 1763–1779.
Roy, J.E., Riesenhuber, M., Poggio, T., Miller, E.K., 2010. Prefrontal cortex activity during flexible categorization. J. Neurosci. Official J. Soc. Neurosci. 30 (25), 8519–8528.
Sawamura, H., Shima, K., Tanji, J., 2002. Numerical representation for action in the parietal cortex of the monkey. Nature 415 (6874), 918–922.
Sohal, V.S., Hasselmo, M.E., 2000. A model for experience-dependent changes in the responses of inferotemporal neurons. Network 11 (3), 169–190.
Steinmetz, M.A., Constantinidis, C., 1995. Neurophysiological evidence for a role of posterior parietal cortex in redirecting visual attention. Cereb. Cortex 5, 448–456.
Swaminathan, S.K., Freedman, D.J., 2012. Preferential encoding of visual categories in parietal cortex compared with prefrontal cortex. Nat. Neurosci. 15, 315–320.
Tamura, K., Takeda, M., Setsuie, R., Tsubota, T., Hirabayashi, T., Miyamoto, K., Miyashita, Y., 2017. Conversion of object identity to object-general semantic value in the primate temporal cortex. Science 357 (6352), 687–692.
Tyulmankov, D., Yang, G.R., Abbott, L.F., 2022. Meta-learning synaptic plasticity and memory addressing for continual familiarity detection. Neuron 110 (3), 544–557 e8.
Uhrig, L., Dehaene, S., Jarraya, B., 2014. A hierarchy of responses to auditory regularities in the macaque brain. J. Neurosci. Official J. Society Neurosci. 34 (4), 1127–1132.
Verguts, T., Fias, W., 2004. Representation of number in animals and humans: a neural model. J. Cogn. Neurosci. 16 (9), 1493–1504.
Viswanathan, P., Nieder, A., 2013. Neuronal correlates of a visual "sense of number" in primate parietal and prefrontal cortices. Proc. Natl. Acad. Sci. 110 (27), 11187–11192.
Viswanathan, P., Nieder, A., 2015. Differential impact of behavioral relevance on quantity coding in primate frontal and parietal neurons. Curr. Biol. CB 25 (10), 1259–1269.

Viswanathan, P., Nieder, A., 2020. Spatial neuronal integration supports a global representation of visual numerosity in primate association cortices. J. Cogn. Neurosci. 32 (6), 1184–1197.
Viswanathan, P., Stein, A.M., Nieder, A., 2024. Sequential neuronal processing of number values, abstract decision, and action in the primate prefrontal cortex. PLoS Biol. 22 (2), e3002520.
Vogels, R., Sary, G., Orban, G.A., 1995. How task-related are the responses of inferior temporal neurons? Vis. Neurosci. 12, 207–214.
Wang, L., Uhrig, L., Jarraya, B., Dehaene, S., 2015. Representation of numerical and sequential patterns in macaque and human brains. Curr. Biol. CB 25 (15), 1966–1974.
Winters, B.D., Saksida, L.M., Bussey, T.J., 2008. Object recognition memory: neurobiological mechanisms of encoding, consolidation and retrieval. Neurosci. Biobehav. Rev. 32 (5), 1055–1070.
Woloszyn, L., Sheinberg, D.L., 2012. Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. Neuron 74 (1), 193–205.
Wu, Z., Buckley, M.J., 2022. Prefrontal and medial temporal lobe cortical contributions to visual short-term memory. J. Cogn. Neurosci. 35 (1), 27–43.
Xiang, J.Z., Brown, M.W., 1998. Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. Neuropharmacology 37 (4-5), 657–676.
Zhang, T., Britten, K.H., 2010. The responses of VIP neurons are sufficiently sensitive to support heading judgments. J. Neurophysiol. 103 (4), 1865–1873.
Zhang, T., Britten, K.H., 2011. Parietal area VIP causally influences heading perception during pursuit eye movements. J. Neurosci. Official J. Society Neurosci. 31 (7), 2569–2575.
Zhou, Y., Rosen, M.C., Swaminathan, S.K., Masse, N.Y., Zhu, O., Freedman, D.J., 2021. Distributed functions of prefrontal and parietal cortices during sequential categorical decisions. eLife 10, e58782.
Zhou, Y., Zhu, O., Freedman, D.J., 2023. Posterior parietal cortex plays a causal role in abstract memory-based visual categorical decisions. J. Neurosci. 43 (23), 4315–4328.
Zorzi, M., Priftis, K., Umiltà, C., 2002. Brain damage: neglect disrupts the mental number line. Nature 417, 138–139.

Publication 2: Coordinated parietal–frontal neuronal communication is critical for abstract quantity judgments in primates.

This chapter is identical to the published version of:

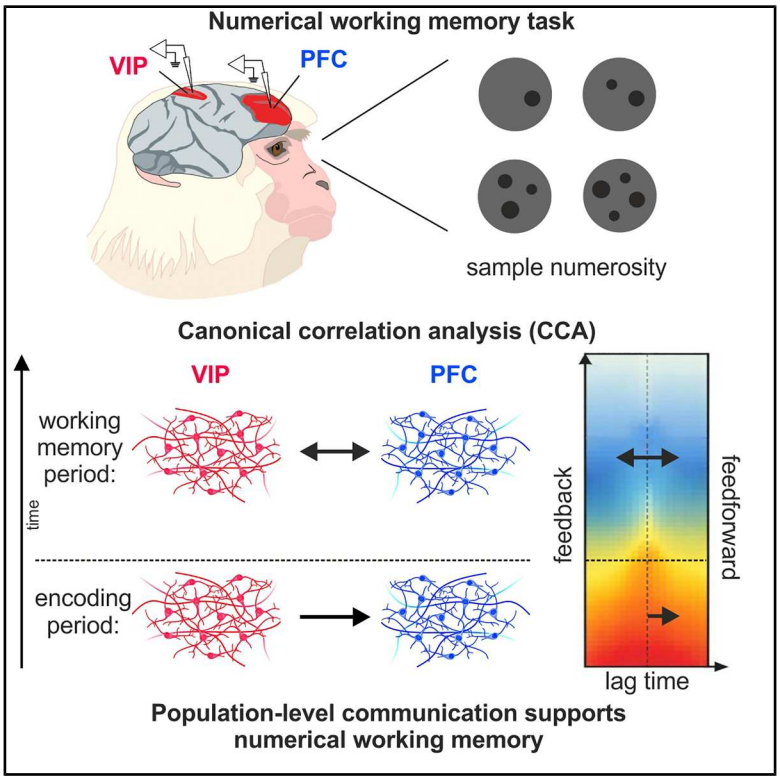
Machts, T., & Nieder, A. (2026). Coordinated parieto-frontal neuronal communication is critical for abstract quantity judgments in primates. *Cell Reports*, *45*(4). https://doi.org/10.1016/j.celrep.2026.117147

Cell Reports

Article

Coordinated parieto-frontal neuronal communication is critical for abstract quantity judgments in primates

Graphical abstract



Highlights

- VIP and PFC show dynamic population coupling during numerosity decisions
- Numerosity-selective neurons drive early feedforward signaling from VIP to PFC
- Sustained bidirectional VIP-PFC interactions support numerical working memory
- Error trials show disrupted parieto-frontal communication late in the delay

Authors

Tobias Machts, Andreas Nieder

Correspondence

andreas.nieder@uni-tuebingen.de

In brief

Using simultaneous population recordings and time-lagged canonical correlation analysis in macaques, Machts and Nieder show that numerosity-selective neurons mediate dynamic VIP-PFC communication. Early feedforward signaling supports encoding, whereas sustained bidirectional interactions maintain numerical information. Disrupted parieto-frontal coupling predicts decision errors.

Machts & Nieder, 2026, Cell Reports *45*, 117147
April 28, 2026
https://doi.org/10.1016/j.celrep.2026.117147

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Coordinated parieto-frontal neuronal communication is critical for abstract quantity judgments in primates

Tobias Machts[1] and Andreas Nieder[1,2,*]
[1]Animal Physiology, Institute of Neurobiology, University of Tuebingen, Auf der Morgenstelle 28, 72076 Tuebingen, Germany
[2]Lead contact
*Correspondence: andreas.nieder@uni-tuebingen.de
https://doi.org/10.1016/j.celrep.2026.117147

SUMMARY

Accurate numerical cognition relies on representing and comparing quantities, a fundamental skill for adaptive behavior. In nonhuman primates, the parieto-frontal network, including the ventral intraparietal area (VIP) and prefrontal cortex (PFC), supports this process, but the functional dynamics and directionality of communication between these regions remain unclear. We examine population-level interactions of simultaneously recorded neuronal activity between VIP and PFC in two male macaques performing a numerosity task. Using time-lagged canonical correlation analyses, we show that correct trials exhibit sustained correlations driven by numerosity-selective neurons, with early feedforward dominance from VIP to PFC following sample onset. By contrast, error trials show weaker correlations, reduced VIP-to-PFC feedforward signaling, and transient breakdowns during the late working memory period. These findings show that coordinated parieto-frontal population activity enables accurate numerical judgments, whereas disrupted interactions impair performance, highlighting the critical role of dynamic, task-dependent interareal communication in categorical decisions.

INTRODUCTION

Humans and nonhuman animals alike rely on the ability to perceive, represent, and manipulate numerical and quantitative information to navigate their environment. Quantitative categories, such as the number of items in a set, its numerosity, enable essential cognitive functions including foraging, social interactions, and decision-making.[1] The capacity to discriminate and compare quantities underlies more complex mathematical reasoning and problem-solving, making numerical cognition a fundamental aspect of adaptive behavior.[2,3]

The parieto-frontal network, encompassing regions such as the intraparietal sulcus (IPS) in the posterior parietal cortex and prefrontal cortex (PFC), plays a pivotal role in processing numerical information.[4–11] In macaques, neurons within the ventral intraparietal area (VIP) and PFC exhibit tuning to visual numerosity, suggesting a direct encoding of quantity.[12–18] This information subsequently, with a temporal delay, appears in the PFC, where it is integrated with other cognitive processes to support working memory and guide decision-making.[19–22]

Despite known anatomical connections between posterior parietal cortex and PFC,[23,24] the functional connectivity between VIP and PFC, as well as the directionality of information flow, has not been directly tested. The fronto-parietal network provides a flexible communication channel supporting attention, working memory, and decision-making. Previous studies using synchronization measures (e.g., coherence, phase locking, correlation dynamics) have shown that coordinated activity between prefrontal and parietal regions enables integration of sensory evidence with executive control, supporting both bottom-up and top-down processes.[25–30] If parieto-frontal communication is critical for numerical cognition, these interactions should be robust during correct delayed judgments and break down during errors.

We tested and extended this hypothesis in monkeys performing a delayed number-matching task by examining simultaneously recorded parieto-frontal neuron populations using time-lagged canonical correlation analysis (CCA), a dimensionality reduction technique that identifies shared population activity patterns between two brain areas.[31] CCA extracts linear combinations of neuronal activity—canonical dimensions—that are maximally correlated across areas. Unlike traditional synchronization measures, it identifies the specific population subspaces mediating inter-areal communication and, by systematically introducing temporal lags, reveals the structure and directionality of interactions: feedforward signaling occurs when activity in an upstream area precedes correlated activity in a downstream area, and feedback occurs in the reverse direction. We used this approach to investigate how VIP and PFC communicate during numerical judgments, providing new insights into the neural mechanisms of numerical cognition and highlighting the critical role of intact parieto-frontal communication for cognitive performance.

RESULTS

Two rhesus monkeys performed a delayed match-to-sample task in which they matched the number of dots in visual arrays

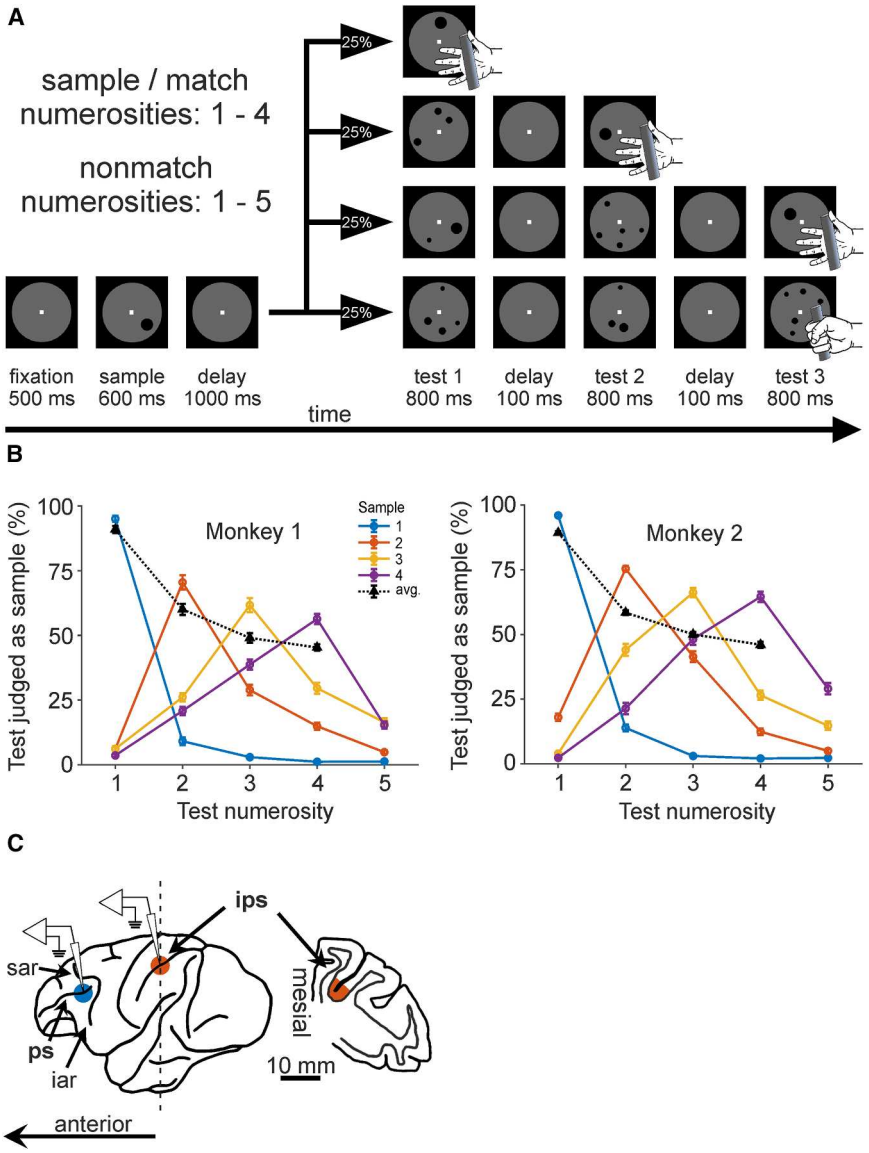


Figure 1. Task protocol, behavioral performance, and recording sites

(A) Monkeys performed a delayed match-to-sample task with dot arrays. Each trial began when the monkey grasped a response bar and maintained fixation on a central target. Sample numerosities (1–4) were presented, followed by a delay period. During up to three sequential test phases (test1–test3), the monkey released the bar for matching numerosities and withheld responding for nonmatching numerosities (1–5). Test phases were pseudo-randomly presented, resulting in four trial conditions. Reproduced from Machts et al.[20]

(B) Behavioral performance of the two monkeys for sample numerosities 1–4. Functions show the probability of judging test displays as matching the sample numerosity (four colors, solid lines). Peaks indicate correct matches, while data points in the flanks reflect errors to nonmatching numerosities. The black dotted line shows the average performance across sample numerosities. Error bars (very small) represent SEM across sessions.

(C) Schematic diagram of the lateral macaque brain showing recording sites in PFC (blue) and VIP (red).

IPS, intraparietal sulcus; PS, principal sulcus; SAR, superior arcuate sulcus; IAR, inferior arcuate sulcus.

(Figure 1A).[20] Sample displays contained 1–4 dots, and test displays contained 1–5 dots. To control for low-level visual features, half of the trials used a standard protocol (dots of random size and location) and half used a control protocol (constant overall dot density and area across numerosities). Each trial began when the monkey grasped a response bar and maintained fixation on a central target. Sample numerosities were presented, followed by a delay period. During up to three sequential test phases (test1–test3), the monkey released the bar for matching numerosities and withheld responding for nonmatching numerosities. Test phases were presented pseudo-randomly, resulting in four trial types, and sample numerosities and trial types were balanced across trials.

Behavioral performance

Both monkeys mastered the task, with mean performance across sessions significantly above chance level of 25% (monkey 1: 27 sessions, 61.32 ± 1.36%, $p < 0.001$; monkey 2: 33 sessions, 60.95 ± 0.58%, $p < 0.001$, Binomial test). To assess their ability to identify the correct matching numerosity, behavioral performance functions were computed for each session, reflecting the probability of judging test numerosities as equal to the sample. For all sample numerosities, both monkeys selected the correct numerosity more frequently than nonmatches and exhibited clear numerical size and distance effects indicative of the approximate number system (Figure 1B).[12,32,33]

Canonical correlation analysis

While the monkeys performed the delayed matching task, we simultaneously recorded neuronal activity in dorsolateral PFC (dlPFC) and VIP of the intraparietal sulcus (60 sessions), brain areas known to represent numerical information.[8] Single-neuron numerosity tuning from this dataset has been reported previously.[20] Here, we focused on analyzing network interactions between VIP and PFC populations.

To investigate population-level interactions, we applied CCA, a dimensionality reduction approach extracting activity patterns in one area that are maximally predictive of activity in the other (Figure 2).[29] To enhance data robustness, we analyzed voltage-thresholded multi-unit activity (MUA) recorded simultaneously from multiple electrodes in PFC and VIP. Of the initial 60 sessions,

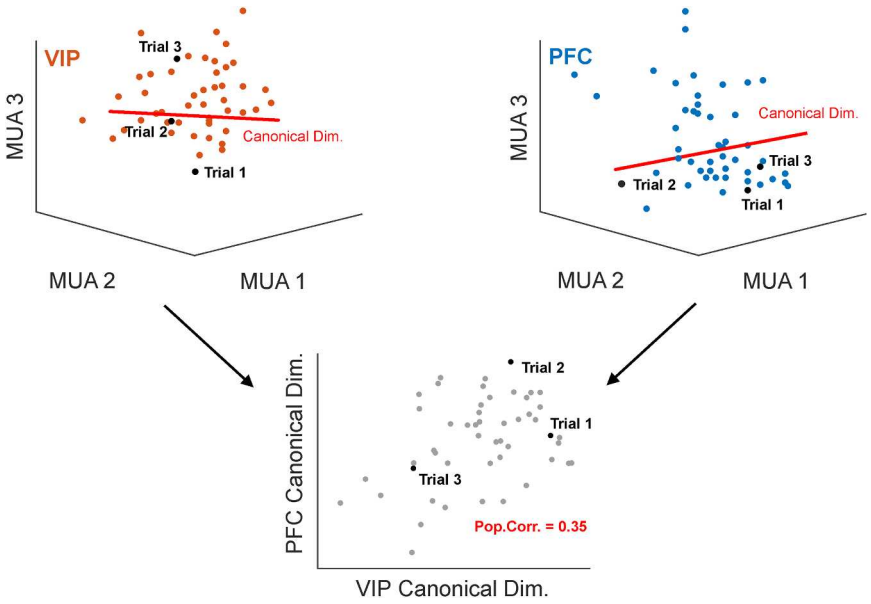


Figure 2. Canonical correlation analysis captures population-level interactions between neurons

In this example, VIP (red, left) and PFC (blue, right) multi-unit activity (MUA) were recorded simultaneously from three electrodes per area. Each circle represents population activity on a single trial; for each VIP activity point, a corresponding PFC point exists. Red axes indicate the first pair of canonical dimensions identified by CCA. Correlating VIP and PFC activity projected onto these dimensions yields a population correlation coefficient (bottom).

After Semedo et al.[29]

we included only those with at least four recording sites per brain area and a minimum of 50 correct and 50 error trials, pooling sessions across numerosities and protocol types (standard vs. control). Using these criteria, 43 sessions were retained for analysis.

To characterize how activity in one area relates to activity in the other across time lags, we applied CCA to population activity vectors from the PFC and VIP, introducing temporal shifts between them.[29] Since these areas are reciprocally connected, with VIP anatomically upstream of PFC, and because prior work shows that neuronal responses and selectivity in VIP lead those in PFC,[17,19,34] we set the maximum time lag to ±40 ms.

We computed a population correlation function (CCA-derived correlation as a function of time lag) across task epochs, from fixation through sample presentation and into the working memory delay. Subsequent Test periods were excluded from analysis due to confounding motor responses. This yielded a two-dimensional population correlation map of the coefficients (Figure 3B).

VIP activity was most strongly correlated with later PFC activity in correct trials

We first examined network interactions based on population correlations during correct trials (Figure 3). Because CCA is sensitive to trial count, we equalized the number of trials across sessions, which allowed us to compute reliable averages (see STAR Methods). Figure 3B shows the average VIP-PFC population correlation map for correct trials.

During the pure fixation period, correlations were absent and did not reach the significance threshold ($p < 0.05$, two-sided; below the 97.5th percentile of the shuffled distribution), indicating chance-level interactions between VIP and PFC. Shortly after sample onset, however, correlation coefficients rose significantly above chance (as indicated by the yellow colors in Figure 3B), with higher values for positive (i.e., VIP leading PFC) than negative lags (right side of the surface plot in Figure 3B). This indicates that VIP activity was most strongly correlated with later PFC activity, consistent with feedforward interaction (until ~200 ms after sample onset). Population correlations remained significant, though lower, until the end of the sample period; however, they did not maintain feedforward dominance during the later phase (400–700 ms). Although correlations declined during the delay, they stayed consistently above chance, indicating ongoing communication between VIP and PFC across time lags throughout the task.

We quantified the lag at which canonical correlations peaked for each time bin using a centroid-based estimation (Figure S1). For each peak, the centroid (center of mass) was computed relative to zero lag to provide a quantitative measure of delays. The centroid was calculated over the range corresponding to the 95th percentile of correlation values around the peak. During the very early sample period, before neurons responded to the stimulus due to their response latency, correlation maxima were near zero lag. Once response latency was exceeded and neurons became active, the maxima shifted systematically toward positive lags, peaking around +10 ms, consistent with feedforward dominance from VIP to dlPFC. At the sample-to-delay transition, correlation maxima returned near zero and remained centered throughout the delay, indicating balanced bidirectional coupling without a consistent lead-lag relationship.

VIP dominance at the onset of the sample

We addressed feedforward and feedback projections functionally by analyzing the direction and structure of correlated population activity across time lags. Feedforward interactions were defined as stronger correlations when VIP led PFC (positive lags), whereas feedback interactions were defined as stronger correlations when PFC led VIP (negative lags). These analyses were performed across different task epochs (fixation, sample, and delay), and first for correct trial conditions.

To quantify VIP-PFC communication directionality, we calculated the area under the positive-lag curve ($AUC_{feedforward}$, Figure 3C) as a measure of VIP-to-PFC feedforward correlation, and the area under the negative-lag curve ($AUC_{feedback}$, Figure 3A) for PFC-to-VIP feedback, using a sliding-window approach at each time step (see STAR Methods). Both

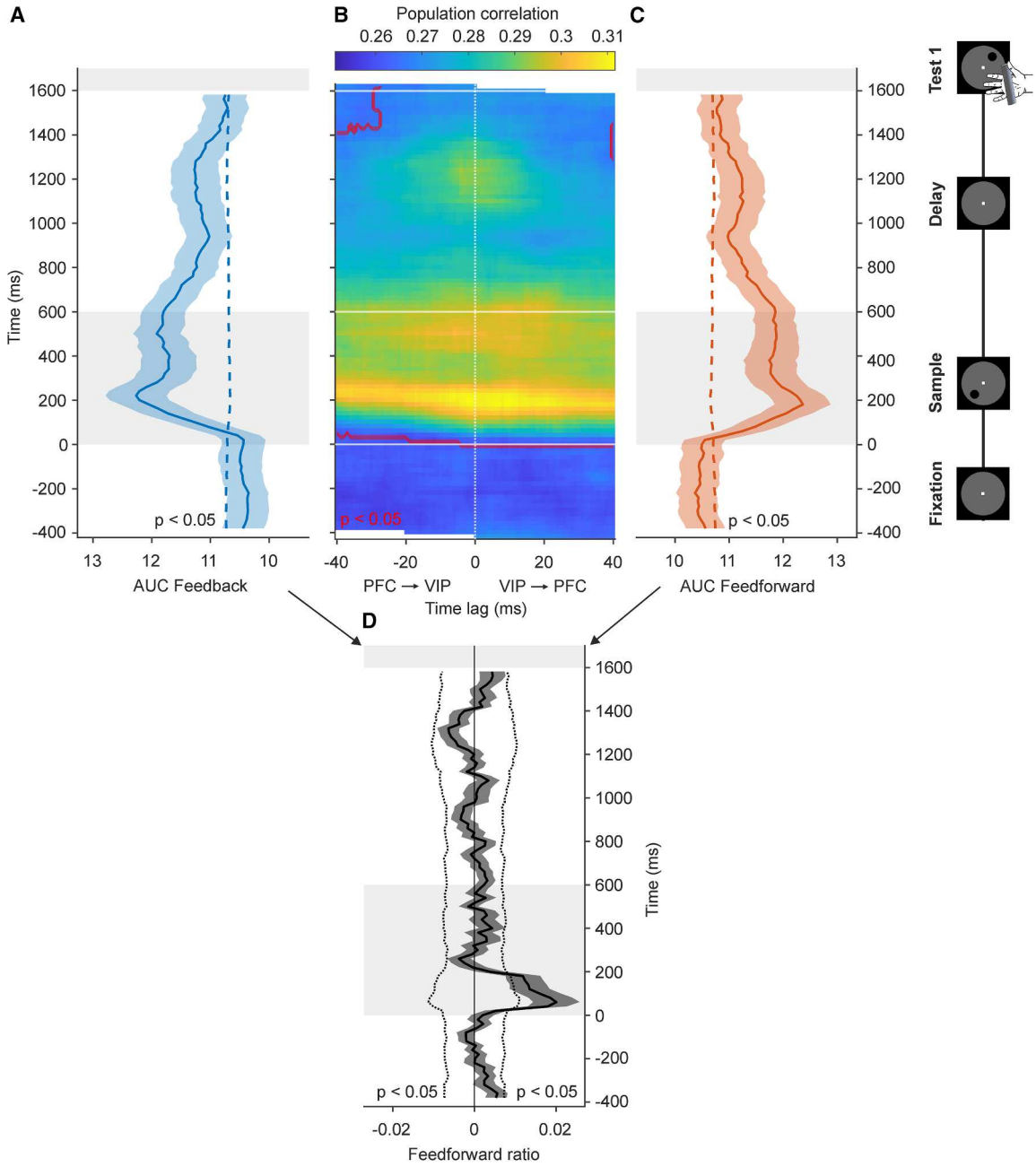


Figure 3. VIP-PFC interactions for correct trials

(A) The area under the feedback sides (AUC, negative lag, PFC → VIP) of the population correlation map (from (B), left half) is shown for correct match trials at all time points. Time 0 is the onset of the sample stimulus, indicated by the gray shaded area. The colored shaded area indicates SEM ($N = 43$ sessions). The dashed horizontal line depicts the significance threshold, which is set to $p < 0.05$ (two-sided).

(B) Averaged population correlation coefficients at all trial times and lag times during correct match trials ($N = 43$ sessions). The horizontal axis represents the time lag between brain areas ($t_1 - t_2$), and the vertical axis represents time relative to sample onset. White horizontal lines indicate the duration of the fixation, sample, delay, and test1 periods. The dashed vertical line indicates zero-lag population correlations ($t_1 = t_2$; see Figure S1A). White areas denote times for which population correlations could not be computed; that is, the PFC activity window had reached either the beginning or the end of the trial. The red boundary indicates the significance threshold, which was set to $p < 0.05$ (two-sided).

(C) Area under the feedforward sides (AUC, positive lag, VIP → PFC) of the population correlation map (from (B), right half) for correct match trials at all time points. Same layout as in (A).

(D) Feedforward ratio at all time points for correct trials. The feedforward ratio is defined as the difference between the areas under the feedforward (positive lag, (A)) and feedback (negative lag, (C)) sides of the population correlation function, divided by their sum. Shaded area indicates SEM ($N = 43$ sessions). The dashed vertical lines indicate the significance threshold set to $p < 0.05$ (two-sided).

feedforward and feedback AUC values remained significantly above chance from sample onset through the end of the delay period.

To quantify directional dominance—whether communication was *more* feedforward or *more* feedback—we contrasted $AUC_{feedforward}$ and $AUC_{feedback}$ and computed a feedforward ratio (Figure 3D). The feedforward ratio is defined as the difference between the area under the positive-lag curve ($AUC_{feedforward}$; Figure 3C) and the area under the negative-lag curve ($AUC_{feedback}$; Figure 3A), divided by their sum at each time step (see STAR Methods).

Significant VIP dominance emerged shortly after sample onset (0.018 ± 0.005; 97.5th percentile: 0.010) and persisted until ~200 ms (0.012 ± 0.004; 97.5th percentile: 0.009), indicating predominantly feedforward communication from VIP to PFC. Beyond this window, neither area showed significant directional dominance during the later sample period or the subsequent delay.

Main interaction patterns were qualitatively consistent across animals. In both monkeys, VIP^{-}dlPFC correlations were strongest during the sample period, showing transient feedforward VIP-to-dlPFC dominance around stimulus onset, followed by balanced bidirectional coupling during the delay (Figure S2). Correlation magnitudes varied, but the temporal structure and directionality were similar across subjects.

Inter-areal communication specific to numerical information

To test whether the observed inter-areal communication was specific to numerical information rather than reflecting nonspecific recurrent or bidirectional dynamics, we repeated the canonical correlation analysis as a function of neuronal numerosity selectivity (Figure 4). Each recording site was categorized as numerosity sample-selective, numerosity delay-selective, or numerosity unselective using a one-way ANOVA with the factor numerosity (1–4; $p < 0.01$).[20] Note that some MUAs were selective during both the sample and delay periods.

Across VIP and dlPFC, the majority of recording sites showed numerosity selectivity during either the sample or delay period (VIP: 151/257; dlPFC: 205/270), whereas a smaller fraction did not exhibit task-related selectivity (VIP: 106/257; dlPFC: 65/270). The distribution of selectivity classes across cortical areas is shown in Figure 4D.

Inter-areal population correlations were driven by numerosity-selective units. Correlations were strongest for sample-selective MUAs (Figure 4A), followed by weaker VIP-dlPFC correlations for delay-selective MUAs. (Figure 4B) In contrast, correlations were virtually absent for non-selective MUAs (Figure 4C). Feedforward directionality for numerosity-selective neuron groups was similar.

To further quantify the contribution of task-selective neuronal populations to inter-areal coordination, we computed mean population correlation values across time lags for each selectivity class (Figure 4E). Sample-selective MUAs showed significantly higher mean population correlation values than non-selective MUAs during sample presentation and the subsequent delay period ($p < 0.05$, Mann-Whitney U test, two-sided). Mean correlation values for delay-selective MUAs were lower overall but did not differ significantly from those of sample-selective MUAs.

Importantly, weaker inter-areal correlations in non-selective MUAs cannot be explained by population size. In VIP, where selectivity groups were similar in number (Figure 4D), non-selective MUAs still showed markedly lower correlations. Delay-selective MUAs also exceeded non-selective correlations despite comparable numbers. Overall, these results indicate that VIP-dlPFC communication is specific to numerical information, rather than reflecting nonspecific recurrent or bidirectional dynamics.

Interaction breakdown in error trials during the late delay period

Next, we analyzed balanced numbers of error trials from the same sessions. Error trials were defined exclusively as incorrect behavioral decisions, namely failures to release the bar on match trials or erroneous releases on non-match trials. Trials with fixation breaks, premature responses, or aborted trials were excluded from all analyses.

Figure 5B shows the average VIP-PFC population correlation map for error trials. As in correct trials, correlations during fixation remained below significance ($p < 0.05$, two-sided; below the 97.5th percentile of the shuffled distribution). Shortly after sample onset, correlation coefficients rose significantly above the 97.5th percentile of the shuffled distribution. Correlations were higher for positive than negative lags, indicating again feedforward dominance (VIP leading PFC) after sample onset. Subsequently, correlations were balanced on the positive and negative lag sides, eliminating clear lag dominance. This balance between positive and negative lags persisted throughout the delay period. Notably, in error trials, population correlations dropped below chance at the end of the delay for both positive and negative lags (Figure 5B), whereas correlations in correct trials remained consistently above the significance threshold ($p < 0.05$; Figure 3B).

Interaction differences between correct and error trials

We examined the relative differences (percentage change) in canonical correlation values between correct and error trials for each monkey (Figure S3). In both animals, correlations were higher for correct trials near the end of the sample period and during the late delay. While the magnitude of these differences varied across monkeys, the temporal profile and direction of correct-error modulation were consistent. These results indicate that the reduction of inter-areal coupling during error trials is a reliable phenomenon across subjects.

Next, we compared overall correlation strength between correct and error trials. We computed the percent difference between the population correlation maps of correct and error trials (Figure 6A). No differences were present in the pure fixation period. However, significant correlation differences emerged for positive time lags at the beginning of the sample period and for all time lags exceeding 10% at the end of the sample and the end of the delay period. Statistical significance was confirmed with a binwise two-sided Wilcoxon signed-rank test, indicated on the map in Figure 6A by a red line ($p < 0.05$).

To complement this analysis, we calculated a mean population correlation function, averaging correlations across all time

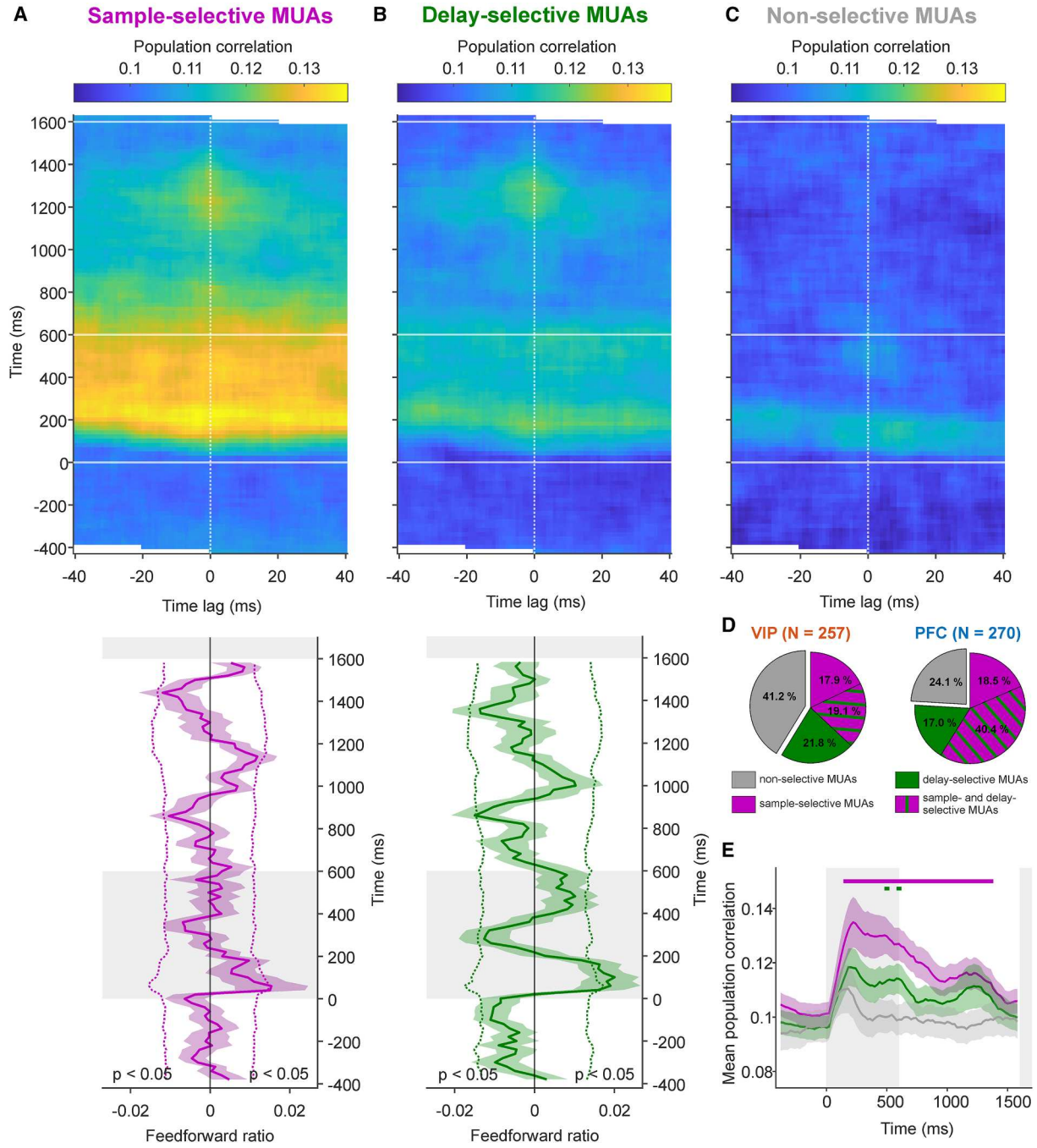


Figure 4. Numerosity-selective neuronal populations contribute to VIP-dlPFC population correlations

(A) Averaged population correlation coefficients across all trial and lag times for sample-selective MUAs. Analyses include correct trials only. Layout matches Figure 3B. The corresponding feedforward ratio at all time points is shown at the bottom.

(B) Averaged population correlation coefficients across all trial and lag times for delay-selective MUAs. Layout as in (A).

(C) Averaged population correlation coefficients across all trial and lag times for non-selective MUAs. Layout as in (A).

(D) Distribution of task-selective MUAs in VIP and dlPFC. Pie charts show the proportion of sample-selective, delay-selective, and non-selective MUAs in VIP (*left*) and dlPFC (*right*). Some MUAs were selective in both periods (dashed shading). Non-selective MUAs showed no significant modulation in either epoch.

(E) Mean time-resolved population correlations for correct trials, shown by selectivity class: sample-selective (magenta), delay-selective (green), and non-selective (gray). Mean correlations represent the average across all time lags between brain areas ($t_1 - t_2$). Time 0 corresponds to sample onset. The magenta bar at the top marks time bins where sample-selective MUAs differ significantly from non-selective MUAs ($p < 0.05$, two-sided Mann-Whitney U test). The green bar at the top marks time bins where delay-selective MUAs differ significantly from non-selective MUAs ($p < 0.05$, two-sided Mann-Whitney U test). Dark shading indicates SEM.

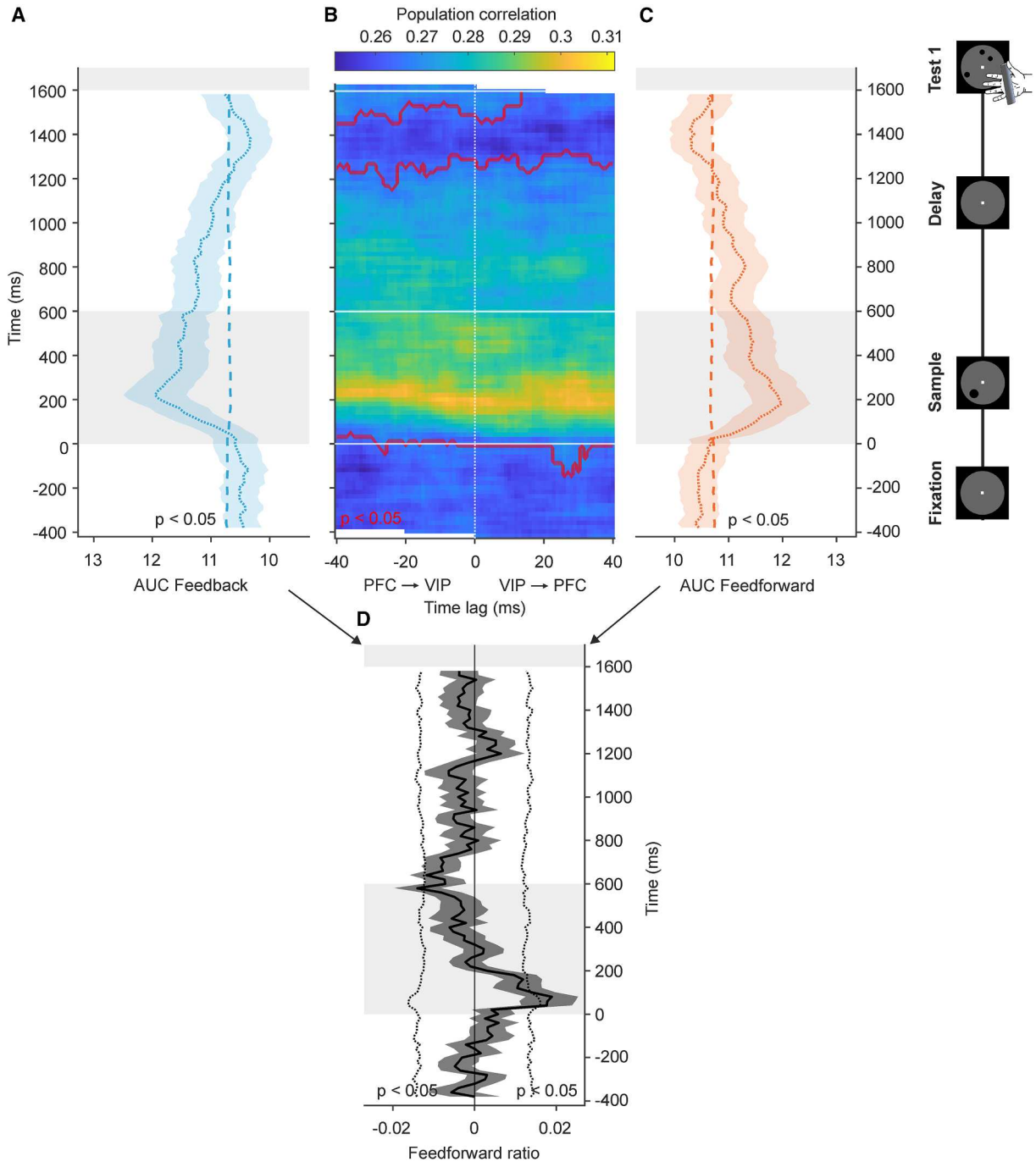


Figure 5. VIP-PFC interactions for error trials

(A) The area under the feedback sides (AUC, negative lag, PFC → VIP) of the population correlation map (from (B)) is shown for error trials at all time points. Time 0 is the onset of the sample stimulus, indicated by the gray shaded area. The colored shaded area indicates SEM ($N = 43$ sessions, same sessions and scaling as in Figure 3). The dashed horizontal line depicts the significance threshold, which is set to $p < 0.05$ (two-sided).

(B) Averaged population correlation coefficients at all trial times and lag times during error trials ($N = 43$ sessions, same sessions and scaling as in Figure 3). The dashed vertical line indicates zero-lag population correlations ($t_1 = t_2$, see Figure S1B). Same layout as in Figure 3B.

(C) The area under the feedforward sides (AUC, positive lag, VIP → PFC) of the population correlation map (from (B)) for error trials at all time points. Same layout as in (A).

(D) Feedforward ratio at all time points for error trials. Same layout as in Figure 3D.

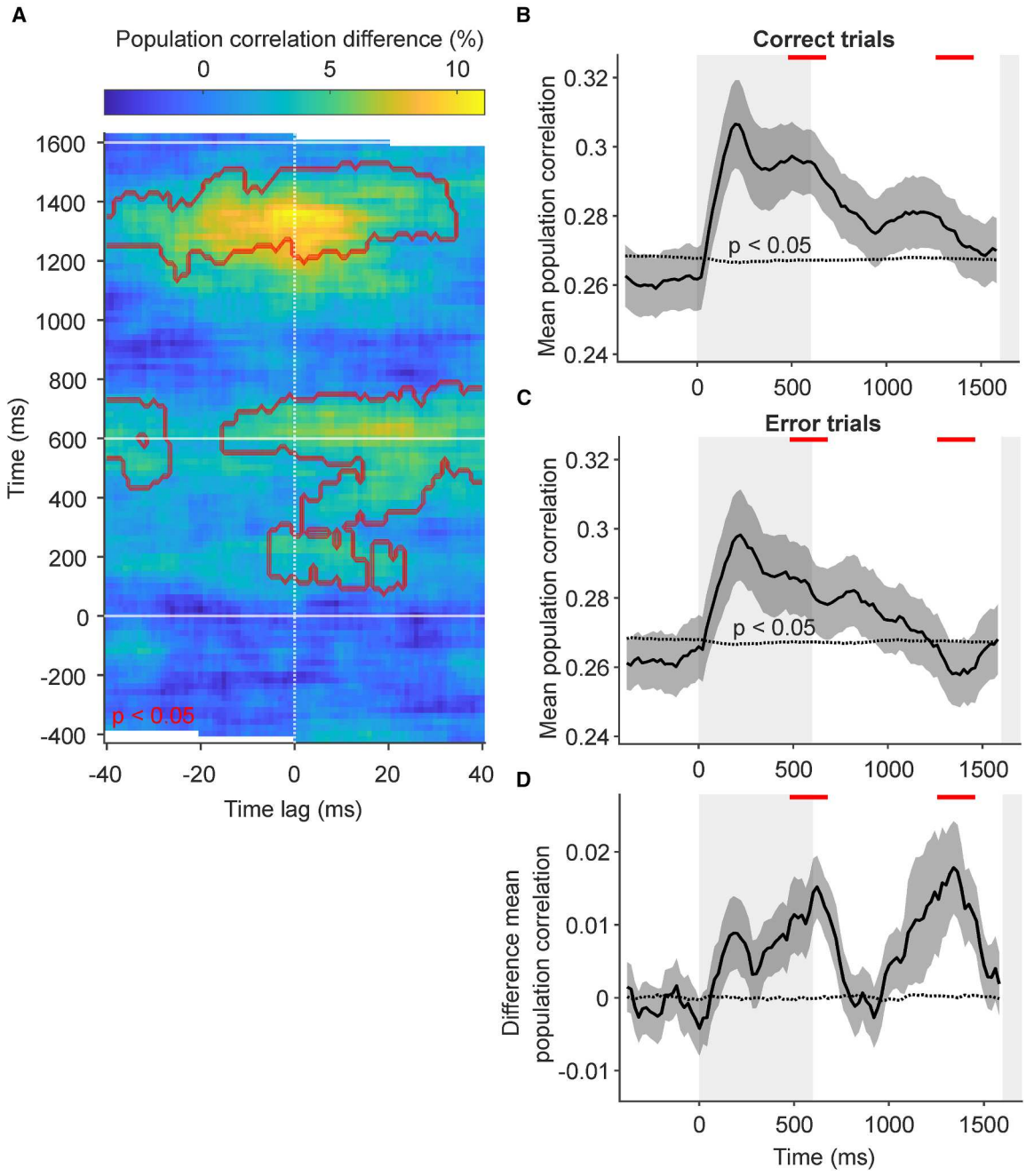


Figure 6. VIP-PFC interaction differences between correct and error trials

(A) The difference in percent between the population correlation coefficient maps of correct and error trials ($N = 43$ sessions, from Figures 3B and 5B). Same layout as in Figure 3B. The red boundary indicates the results of the bin-wise Wilcoxon signed-rank test ($p < 0.05$, two-sided).

(B) Mean time-resolved population correlations for correct trials. The mean population correlation refers to the average of all population correlations across all time lags between brain areas ($t_1 - t_2$). Time 0 is the onset of the sample stimulus indicated by the gray shaded area. The dark shaded areas indicate SEM ($N = 43$ sessions). The dashed horizontal line depicts the significance threshold set to $p < 0.05$ (two-sided). Red bars on the top of the subplot indicate significant differences between the mean correlation coefficients per time bin for error trials ((C), $p < 0.05$, Wilcoxon signed-rank test, two-sided).

(C) Mean population correlations for error trials. Red bars on the top of the subplot indicate significant differences between the mean correlation coefficients per time bin for correct trials ((B), $p < 0.05$, Wilcoxon signed-rank test, two-sided). Same layout as in (B).

(D) The absolute difference between the mean population correlations of correct and error trials ($N = 43$ sessions, from (B) and (C)). The dashed horizontal line depicts the difference between the shuffled distributions of correct and error trials. Same layout as in (B).

lags between brain areas (−40 to +40 ms). These functions were computed separately for correct (Figure 6B) and error trials (Figure 6C) and compared using a binwise two-sided Wilcoxon signed-rank test. Consistent with previous results, correlations were significantly higher in correct trials at the end of the sample presentation (500–680 ms) and during the later delay phase (1260–1440 ms). The difference between the two functions (correct − error) and the significant difference periods are shown in Figure 6D. A similar picture emerged for zero-lag population correlations in correct and error trials (Figure S4). These results indicate that stronger and sustained VIP-PFC population correlations, particularly at the end of the sample and delay periods, are associated with correct numerical judgments, whereas weakened or disrupted correlations characterize error trials.

DISCUSSION

In this study, we examined how neuronal populations in VIP and dlPFC interact during numerical working memory using CCA. Our results show that parieto-frontal communication is dynamic and task-dependent: during stimulus presentation, VIP activity predicted later PFC activity, consistent with a feedforward influence. This early dominance transitioned into balanced bidirectional communication, which persisted across the delay period. Inter-areal population correlations were driven almost exclusively by numerosity-selective neurons, particularly sample-selective units and, to a lesser extent, delay-selective units. In contrast, non-selective MUAs showed negligible correlations across time lags. These results indicate that the reported inter-areal communication is specific to numerical information, rather than reflecting nonspecific recurrent or bidirectional dynamics related to attention or task engagement. Critically, the strength and persistence of these correlations distinguished correct from error trials, suggesting that intact VIP-PFC interactions are necessary for accurate numerical judgments.

Population-level insights with CCA

By using CCA, we were able to reveal population subspaces that mediate VIP-PFC communication, going beyond traditional synchronization measures. Most previous studies of fronto-parietal communication have relied on synchronization measures (e.g., coherence, phase locking, or correlation dynamics) to characterize large-scale coupling between brain regions.[23–28] These approaches demonstrate that oscillatory coupling and spike-field coherence carry task-relevant information and reflect the balance of top-down and bottom-up signaling. However, such measures typically assess coupling at the level of individual electrodes, frequency bands, or spike-field relations, without directly probing the population activity subspaces through which feedforward and feedback signals are exchanged.

By contrast, Semedo et al.[29] introduced a dimensionality reduction framework based on CCA that extracts the population-level patterns of activity in one area that are maximally predictive of activity in another. Crucially, they showed that feedforward and feedback interactions recruit different, largely orthogonal subspaces of population activity, implying that information is routed via distinct channels depending on interaction direction. Applying this method to parieto-frontal circuits extends prior work[35] by moving beyond simple synchronization strength to reveal how population-level communication is structured, temporally directed, and functionally segregated between feedforward and feedback signals.

Feedforward dominance at stimulus encoding

The transient feedforward dominance from VIP to PFC shortly after sample onset aligns with anatomical and physiological evidence that the intraparietal sulcus provides an initial source of numerical information to prefrontal cortex.[17] The feedforward lags are consistent with physiological conduction delays. The ∼10 ms VIP response lead over dlPFC after presentation of numerosity stimuli aligns with our previous findings of +11 ms (Figure 9 in Merten et al.[32]) and +13 ms (Figure S6 in Jannasch et al.[33]). Similar feedforward dynamics have been observed for perceptual and attentional signals in parieto-frontal networks.[23,26] Our findings suggest that numerical information is initially extracted in parietal cortex and transmitted to PFC for further evaluation and integration into working memory representations.

Recurrent parieto-frontal loops during maintenance

After the initial feedforward phase, correlations remained significant but showed no consistent dominance of VIP or PFC. This balance suggests that working memory relies on recurrent communication between parietal and prefrontal cortex rather than exclusively on PFC maintenance mechanisms.[19] This view is supported by models of distributed working memory, which propose that prefrontal activity is stabilized and reinforced through reciprocal interactions with posterior cortical regions.[28,36–38] Our findings provide direct evidence at the population level that recurrent VIP-PFC interactions support the maintenance of numerical information over memory delays.

Breakdown of communication in error trials

A key finding is that error trials were associated with a breakdown of VIP-PFC correlations at the end of the delay, a phase critical for transforming maintained representations into decision signals. This disruption suggests that errors arise not merely from weak tuning within areas but from failures in interareal coordination. Similar communication failures have been reported in parieto-frontal networks during attentional lapses[27] and impaired decision-making,[24] pointing to a general mechanism whereby weakened long-range interactions undermine cognitive control and task performance.

Implications for numerical cognition and dyscalculia

These results extend current models of numerical cognition by demonstrating that accurate number judgments require not only numerosity-selective neurons but also intact communication between parietal and prefrontal populations. The observed breakdown of communication in error trials suggests that numerical deficits, whether transient or chronic, may partly reflect impaired fronto-parietal interactions.

Our findings of disrupted or weakened VIP-PFC interactions during errors resonate with evidence from children with developmental dyscalculia, where atypical parieto-frontal connectivity has been consistently reported. Structural and functional

imaging studies show that children with dyscalculia exhibit altered connectivity within numerical brain networks, including both hyperconnectivity during arithmetic tasks[39] and abnormal anatomical graph-theoretical properties.[40] Importantly, these aberrant connectivity patterns are not static: targeted numerical interventions can normalize functional hyperconnectivity,[41] and more recent work demonstrates altered effective connectivity specifically in dyscalculic children.[42] Longitudinal evidence further suggests that mathematical learning deficits originate from atypical development of a fronto-parietal brain network early in childhood.[43] Taken together, these human data align with our observation that successful task performance depends on strong and sustained VIP-PFC communication, supporting the idea that disrupted fronto-parietal interactions constitute a core neural mechanism underlying both transient errors in primates and persistent numerical learning difficulties in humans.

Current treatments for developmental dyscalculia rely largely on extensive behavioral training, often with limited success. A macaque model could provide a valuable platform for testing pharmacological interventions aimed at strengthening network connectivity. Both the parietal association cortex and the dorsolateral prefrontal cortex (dlPFC) rely heavily on neuromodulatory mechanisms. For example, activation of muscarinic M1 receptors (M1Rs) enhances dlPFC circuit function by modulating neuronal excitability through KCNQ potassium channels. This mechanism supports working memory in primates.[44] Cholinergic modulation likely interacts with other neuromodulatory systems, including dopamine and norepinephrine. These systems also shape dlPFC persistent activity and network stability.[45–47] Although speculative, M1R agonists such as *xanomeline* may enhance VIP-dlPFC communication. The macaque model might be suitable for testing such interventions.

Limitations of the study

A limitation of our study is that time-lagged canonical correlation analysis does not provide direct evidence of causal information flow between VIP and dlPFC. The observed temporal asymmetries are consistent with feedforward VIP-to-dlPFC communication but could in principle also arise from a third area projecting to both regions with different delays. For instance, neurons in the medial temporal lobe of humans, such as parahippocampal area and hippocampus, also contain numerosity-selective neurons.[48–50] While VIP-dlPFC interactions carry numerical information, future studies using causal manipulations, such as targeted inactivation or optogenetic perturbation, are needed to establish the functional influence of VIP on dlPFC.

We note that time-lagged canonical correlation analysis benefits from large simultaneously recorded populations, whereas our recordings included an average of 6.0 ± 0.2 sites in VIP and 6.3 ± 0.2 sites in dlPFC per session (43 sessions total). All analyses were based on multi-unit activity, which pools spiking from multiple neurons per electrode and increases robustness. The relatively modest population size may contribute to the lower canonical correlation values (~0.3) compared to larger-scale tCCA studies (e.g., Semedo et al., 2022).[29] Nevertheless, the temporal structure, task selectivity, and inter-areal patterns were consistent across sessions and animals, supporting the validity of our conclusions.

Future work should test causal links between parieto-frontal communication and numerical behavior using perturbation methods or causal modeling. Extending this work to developmental trajectories will also be critical, as fronto-parietal connectivity supporting numerical cognition emerges gradually during childhood and undergoes refinement with learning.[41] Bridging primate neurophysiology with human neuroimaging could help identify conserved network mechanisms and, in turn, inform interventions for disorders such as developmental dyscalculia. More broadly, our approach illustrates how analyzing population-level communication between higher cortical areas can reveal principles of cognitive control and working memory that generalize beyond numerosity processing.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Andreas Nieder (andreas.nieder@uni-tuebingen.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Processed data are publicly available at https://doi.org/10.6084/m9.figshare.31235425.
- Original analysis code is publicly available and provided at https://doi.org/10.6084/m9.figshare.31235425.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This research was supported by DFG grants NI 618/15-1 and NI 618/17-1 (FOR 5159).

AUTHOR CONTRIBUTIONS

T.M. and A.N. designed the experiment, T.M. recorded the data, T.M. and A.N. analyzed the data, T.M. and A.N. wrote the paper, and A.N. supervised the study.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

The authors used ChatGPT, developed by OpenAI, for language editing assistance in this manuscript. ChatGPT was solely utilized for language proofreading purposes and did not contribute to or influence any intellectual content within the manuscript.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND PARTICIPANT DETAILS
  - Subjects
- METHOD DETAILS
  - Experimental apparatus and stimuli

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2026.117147.

Received: October 10, 2025
Revised: February 3, 2026
Accepted: February 27, 2026

REFERENCES

1. Nieder, A. (2020). The Adaptive Value of Numerical Competence. Trends Ecol. Evol. *35*, 605–617.
2. Nieder, A. (2025). The calculating brain. Physiol. Rev. *105*, 267–314.
3. Nieder, A. (2026). The making of number: from content to representation. Trends Cogn. Sci. *S1364–6613*, 00360-2. https://doi.org/10.1016/j.tics.2025.12.011.
4. Sawamura, H., Shima, K., and Tanji, J. (2002). Numerical representation for action in the parietal cortex of the monkey. Nature *415*, 918–922.
5. Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. Cogn. Neuropsychol. *20*, 487–506.
6. Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. Neuron *44*, 547–555.
7. Nieder, A. (2004). The number domain- can we count on parietal cortex? Neuron *44*, 407–409.
8. Nieder, A., and Dehaene, S. (2009). Representation of number in the brain. Annu. Rev. Neurosci. *32*, 185–208.
9. Nieder, A. (2016). The neuronal code for number. Nat. Rev. Neurosci. *17*, 366–382.
10. Harvey, B.M., Ferri, S., and Orban, G.A. (2017). Comparing Parietal Quantity-Processing Mechanisms between Humans and Macaques. Trends Cogn. Sci. *21*, 779–793.
11. Tsouli, A., Harvey, B.M., Hofstetter, S., Cai, Y., van der Smagt, M.J., Te Pas, S.F., and Dumoulin, S.O. (2022). The role of neural tuning in quantity perception. Trends Cogn. Sci. *26*, 11–24.
12. Nieder, A., Freedman, D.J., and Miller, E.K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. Science *297*, 1708–1711.
13. Nieder, A., and Miller, E.K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. Neuron *37*, 149–157.
14. Viswanathan, P., and Nieder, A. (2015). Differential impact of behavioral relevance on quantity coding in primate frontal and parietal neurons. Curr. Biol. *25*, 1259–1269.
15. Eiselt, A.K., and Nieder, A. (2016). Single-cell coding of sensory, spatial and numerical magnitudes in primate prefrontal, premotor and cingulate motor cortices. Exp. Brain Res. *234*, 241–254.
16. Ramirez-Cardenas, A., Moskaleva, M., and Nieder, A. (2016). Neuronal Representation of Numerosity Zero in the Primate Parieto-Frontal Number Network. Curr. Biol. *26*, 1285–1294.
17. Viswanathan, P., and Nieder, A. (2020). Spatial Neuronal Integration Supports a Global Representation of Visual Numerosity in Primate Association Cortices. J. Cogn. Neurosci. *32*, 1184–1197.
18. Viswanathan, P., Stein, A.M., and Nieder, A. (2024). Sequential neuronal processing of number values, abstract decision, and action in the primate prefrontal cortex. PLoS Biol. *22*, e3002520.
19. Nieder, A., and Miller, E.K. (2004). A parieto-frontal network for visual numerical information in the monkey. Proc. Natl. Acad. Sci. USA *101*, 7457–7462.
20. Viswanathan, P., and Nieder, A. (2013). Neuronal correlates of a visual "sense of number" in primate parietal and prefrontal cortices. Proc. Natl. Acad. Sci. USA *110*, 11187–11192.
21. Jacob, S.N., and Nieder, A. (2014). Complementary roles for primate frontal and parietal cortex in guarding working memory from distractor stimuli. Neuron *83*, 226–237.
22. Machts, T., Grüb, J., and Nieder, A. (2025). Neuronal encoding of recognition memory for numerical quantities in macaque intraparietal and prefrontal cortices. Prog. Neurobiol. *251*, 102794.
23. Chafee, M.V., and Goldman-Rakic, P.S. (2000). Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. J. Neurophysiol. *83*, 1550–1566.
24. Oguchi, M., and Sakagami, M. (2022). Dissecting the Prefrontal Network With Pathway-Selective Manipulation in the Macaque Brain-A Review. Front. Neurosci. *16*, 917407.
25. Buschman, T.J., and Miller, E.K. (2007). Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. Science *315*, 1860–1862.
26. Salazar, R.F., Dotson, N.M., Bressler, S.L., and Gray, C.M. (2012). Content-specific fronto-parietal synchronization during visual working memory. Science *338*, 1097–1100.
27. Pesaran, B., Nelson, M.J., and Andersen, R.A. (2008). Free choice activates a decision circuit between frontal and parietal cortex. Nature *453*, 406–409.
28. Gregoriou, G.G., Gotts, S.J., Zhou, H., and Desimone, R. (2009). High-frequency, long-range coupling between prefrontal and visual cortex during attention. Science *324*, 1207–1210.
29. Dotson, N.M., Salazar, R.F., and Gray, C.M. (2014). Frontoparietal correlation dynamics reveal interplay between integration and segregation during visual working memory. J. Neurosci. *34*, 13600–13613.
30. Crowe, D.A., Goodwin, S.J., Blackman, R.K., Sakellaridi, S., Sponheim, S.R., MacDonald, A.W., 3rd, and Chafee, M.V. (2013). Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. Nat. Neurosci. *16*, 1484–1491.
31. Semedo, J.D., Jasper, A.I., Zandvakili, A., Krishna, A., Aschner, A., Machens, C.K., Kohn, A., and Yu, B.M. (2022). Feedforward and feedback interactions between visual cortical areas use different population activity patterns. Nat. Commun. *13*, 1099.
32. Merten, K., and Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. J. Cogn. Neurosci. *21*, 333–346.
33. Jannasch, L.L., Grüb, J., and Nieder, A. (2025). Shared cognitive biases influence numerical judgments in macaques and crows. Proc. Natl. Acad. Sci. USA *122*, e2512219122.
34. Tudusciuc, O., and Nieder, A. (2009). Contributions of primate prefrontal and posterior parietal cortices to length and numerosity representation. J. Neurophysiol. *101*, 2984–2994.
35. Jacob, S.N., Hähnke, D., and Nieder, A. (2018). Structuring of Abstract Working Memory Content by Fronto-parietal Synchrony in Primate Cortex. Neuron *99*, 588–597.e5.
36. Merchant, H., Crowe, D.A., Robertson, M.S., Fortes, A.F., and Georgopoulos, A.P. (2011). Top-down spatial categorization signal from prefrontal to posterior parietal cortex in the primate. Front. Syst. Neurosci. *5*, 69.
37. Christophel, T.B., Klink, P.C., Spitzer, B., Roelfsema, P.R., and Haynes, J.D. (2017). The Distributed Nature of Working Memory. Trends Cogn. Sci. *21*, 111–124.

38. Genovesio, A., Wise, S.P., and Passingham, R.E. (2014). Prefrontal-parietal function: from foraging to foresight. Trends Cogn. Sci. *18*, 72–81.

39. Rosenberg-Lee, M., Ashkenazi, S., Chen, T., Young, C.B., Geary, D.C., and Menon, V. (2015). Brain hyper-connectivity and operation-specific deficits during arithmetic problem solving in children with developmental dyscalculia. Dev. Sci. *18*, 351–372.

40. Estévez-Pérez, N., Sanabria-Díaz, G., Castro-Cañizares, D., Reigosa-Crespo, V., and Melie-García, L. (2023). Anatomical connectivity in children with developmental dyscalculia: A graph theory study. Prog. Brain Res. *282*, 17–47.

41. Michels, L., O'Gorman, R., and Kucian, K. (2018). Functional hyperconnectivity vanishes in children with developmental dyscalculia after numerical intervention. Dev. Cogn. Neurosci. *30*, 291–303.

42. Schwizer Ashkenazi, S., McCaskey, U., O'Gorman Tuura, R., and Kucian, K. (2025). Altered Effective Connectivity of the Numerical Brain in Children With Developmental Dyscalculia. J. Neurosci. Res. *103*, e70066.

43. Kuhl, U., Sobotta, S., and Legascreen Consortium; and Skeide, M.A. (2021). Mathematical learning deficits originate in early childhood from atypical development of a frontoparietal brain network. PLoS Biol. *19*, e3001407.

44. Galvin, V.C., Yang, S.T., Paspalas, C.D., Yang, Y., Jin, L.E., Datta, D., Morozov, Y.M., Lightbourne, T.C., Lowet, A.S., Rakic, P., et al. (2020). Muscarinic M1 Receptors Modulate Working Memory Performance and Activity via KCNQ Potassium Channels in the Primate Prefrontal Cortex. Neuron *106*, 649–661.e4.

45. Goldman-Rakic, P.S., Muly, E.C., 3rd, and Williams, G.V. (2000). D(1) receptors in prefrontal cells and circuits. Brain Res. Rev. *31*, 295–301.

46. Arnsten, A.F.T. (2011). Catecholamine influences on dorsolateral prefrontal cortical networks. Biol. Psychiatry *69*, e89–e99.

47. Ott, T., and Nieder, A. (2019). Dopamine and Cognitive Control in Prefrontal Cortex. Trends Cogn. Sci. *23*, 213–234.

48. Kutter, E.F., Bostroem, J., Elger, C.E., Mormann, F., and Nieder, A. (2018). Single Neurons in the Human Brain Encode Numbers. Neuron *100*, 753–761.e4.

49. Kutter, E.F., Dehnen, G., Borger, V., Surges, R., Mormann, F., and Nieder, A. (2023). Distinct neuronal representation of small and large numbers in the human medial temporal lobe. Nat. Hum. Behav. *7*, 1998–2007.

50. Kutter, E.F., Dehnen, G., Borger, V., Surges, R., Nieder, A., and Mormann, F. (2024). Single-neuron representation of nonsymbolic and symbolic number zero in the human medial temporal lobe. Curr. Biol. *34*, 4794–4802.e3.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental models: Organisms/strains		
Macaca mulatta	German Primate Center, Göttingen	www.dpz.eu
Software and algorithms		
NIMH Cortex	National Institute of Mental Health	https://www.nimh.nih.gov/labs-at-nimh/research-areas/clinics-and-labs/ln/shn/software-projects.shtml
MAP Data Acquisition System	Plexon	https://plexon.com/
MATLAB R2020b	MathWorks	https://www.mathworks.com
MATLAB code	MathWorks	https://doi.org/10.6084/m9.figshare.31235425

EXPERIMENTAL MODEL AND PARTICIPANT DETAILS

Subjects

Two male adult rhesus monkeys (*Macaca mulatta*) were implanted with two recording chambers each, centered over the principal sulcus in the dorsolateral prefrontal cortex (PFC) and the intraparietal sulcus (IPS) in the posterior parietal cortex (monkey 1 and monkey 2, both left hemisphere). Chamber implantation was guided by anatomical magnetic resonance imaging (MRI) and stereotaxic measurements. All surgeries were performed under sterile conditions while the monkeys were under general anesthesia. The monkeys received postoperative antibiotics and analgesics. All procedures were performed in accordance with the guidelines for animal experimentation. All experimental procedures had been ethically reviewed and approved and were regulated and supervised by the national authorities, the Regierungspräsidium Tübingen, Germany.

METHOD DETAILS

Experimental apparatus and stimuli

Monkeys were seated inside primate chairs in a chamber 57 cm away from a 15″ flat screen monitor (with a resolution of 1,024 by 768 pixels and a refresh rate of 75 Hz). The NIMH Cortex software (Bethesda, MD) was used to present the stimuli and behavioral data acquisition. The monkeys' eye gaze was tracked by an infrared eye tracking system (ISCAN, Cambridge, MA).

We used random dot arrays as numerosity stimuli ranging between 1 and 5 dots (diameter range from 0.5° to 0.9° of visual angle). Each unique dot pattern was presented on a gray background circle subtending 5.4° of visual angle and was used as sample and test stimulus. A custom-written MATLAB (The Mathworks, Natick, MA) program generated the stimuli anew before each recording session to prevent the monkeys from memorizing stimulus sequences. Sample and test stimulus of the same countable numerosity were never identical within a trial and session. In order to control for low-level, non-numerical visual features in half of the trials, dot diameters were selected randomly (standard protocol). In the other half of the trials, dot density and total occupied area were equated across stimuli (control protocol). Trials of the standard and control protocol were pseudo-randomly and unpredictably presented.

Behavioral protocol

We trained two macaque monkeys on a modified version of the delayed match-to-numerosity task in which the monkeys had to match test numerosities to sample numerosities (see Machts et al.[20]). Nonmatching numerosities that were presented between sample and matching test stimulus had to be ignored. Up to three sequential numerical nonmatches were presented, i.e., the task consisted of up to three test phases and in each test phase the right or wrong numerosity could occur. A trial was initiated from the monkeys by grabbing a bar and maintaining eye fixation within 3.5° of visual angle of a central white dot. Trials were immediately aborted (blue screen appears) and excluded from further analysis if the animals broke fixation. At first, a gray circular background was shown for 500 ms (pure fixation period). Subsequently, the sample stimulus containing 1 to 4 dots appeared for 600 ms. This sample stimulus had to be memorized over a delay of 1000 ms (again gray background circle) and had to be compared in a following sequence of test numerosities. Each test stimulus was presented for 800 ms with a 100 ms delay between each other. For the case that the numerosity in test stimulus was same as in the sample (match situation, 1–4), monkeys had to respond with a bar release. If the numerosities were different (nonmatch situation, 1–5), the animals continued to hold the bar. A nonmatch numerosity was never shown twice per trial. Sample numerosities and trial types (i.e., if no nonmatch, one nonmatch, two nonmatch or three nonmatch numerosities were

presented) were balanced across conditions and displayed with equal probability (25% each, pseudo-randomly intermixed). Thus, monkeys could not predict when or whether the match stimulus would appear. Correct decision per trial was rewarded with a drop of water. In 25% of the trials, when presenting three nonmatch numerosities in sequence, monkeys were rewarded when holding the bar throughout the whole trial until display-offset. Since the number of trials a monkey can perform during an experimental session is limited (on average monkey 1 completed 483 trials and monkey 2 534 trials), not all possible combinations of sample (1–4) and test numerosities (1–5) could be presented. Therefore, a predefined collection of 64 unique sample-test-combinations was shown.

Neurophysiological recordings

In each session, arrays of up to eight glass-coated 1 MΩ tungsten microelectrodes (Alpha Omega LTD, Israel) were inserted in each recording chamber using custom-made screw microdrives and a grid (Crist Instruments, USA) with 1 mm spacing. To access the IPS, electrodes were passed along the course of it to a depth of 9–10 mm. A MAP Plexon system was used for signal acquisition, amplification, filtering and digitalization. Multi-unit selection was performed offline by adjusting a threshold that separates noise from neuronal activity (Plexon Systems, USA).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis

To focus on network interactions, we analyzed thresholded multi-unit activity (MUA) rather than single-unit activity, using one MUA trace per electrode (up to eight per area). Across 43 sessions, an average of 6.0 ± 0.2 sites were recorded per session in VIP and 6.3 ± 0.2 in dlPFC. Using MUA increases robustness by pooling spikes from multiple neurons per electrode and allowed us to correlate the full population activity between areas for all trials. Electrodes with an average firing rate below 0.5 spikes/s across trials were excluded from subsequent analyses.

All values in the main text and figures refer to the mean ± SEM. SEM was calculated as the standard deviation divided by the square root of the number of samples.

To ensure that our analysis was not contaminated by motor activity, we analyzed the network interactions between PFC and VIP just until the end of the delay period, i.e., the following analyses considered the fixation (500 ms), sample (600 ms) and delay period (1000 ms). We set a criterion of at least 50 recorded trials per correct and error trials, respectively. Furthermore, we used just sessions with at least four intact electrodes per brain area in order to capture robust brain interactions. 43 of 60 sessions fulfilled this criterion (12 sessions from monkey 1, 31 sessions from monkey 2).

Population correlation analysis

The population correlation analysis closely follows the description provided by Semedo et al.[29] However, instead of single-unit activity, we correlated the entire recorded neuronal activity (multi-unit activity) of the inserted electrodes (freed from noise). To capture the relationship between neuronal activity across the two brain areas, we filtered out the 'mean signal' across the trials. To do this, we first z-scored the multi-unit activity and then subtracted the corresponding peri-stimulus time histogram (PSTH), resulting in residual activity. This process was repeated for each electrode separately, without considering the different sample sizes and protocols as individual groups. We then used CCA to identify pairs of dimensions in each brain area such that the correlation between the projected activity onto these dimensions is maximized. This can be described by the following formula:

$$\underset{\boldsymbol{a},\boldsymbol{b}}{\operatorname{argmax}} corr(\boldsymbol{Xa}, \boldsymbol{Yb}),$$

where X is a $n \times p_x$ matrix containing the residual activity of the VIP population, Y is a $n \times p_y$ matrix containing the residual activity of the PFC population, n represents the number of data points, and p_x and p_y are the number of electrodes in each of the two areas. The vectors a and b are one-dimensional and define dimensions in the population activity space of each area. CCA can identify additional pairs of dimensions by searching for vectors that maximize the same correlation, provided they are uncorrelated with the initial pair of canonical variables. However, our CCA identified just one pair of dimensions with significant population correlations, since the correlation values of the second canonical pair were, on average across all tested sessions, around 80% lower (correct trials: 78.1 ± 2.26%; error trials: 76.9 ± 2.45%; N = 43 sessions). Therefore, we only considered the first canonical pair for our analyses.

To measure population correlations at different time points, we defined two time windows with a length of 160 ms in each area. These time windows with the starting times t_1 and t_2 relative to trial onset were slide in 20 ms steps along the trial length. While doing this we additionally shifted both windows against each other with a maximal lag of ±40 ms in 1 ms steps ($d = t_1 - t_2$) to compute population correlations at different lag periods. Thus, defining the time within the trial as $t = t_1$, each entry in the population correlation can described as follows:

$$C(t, d) = P(t, t + d) = \underset{\boldsymbol{a},\boldsymbol{b}}{\max} corr(\boldsymbol{X}_t\boldsymbol{a}, \boldsymbol{Y}_{t+d}\boldsymbol{b}),$$

Due to the relatively small number of trials per recording session (at least 50 trials per group; see above), we counted the spikes in each time window in non-overlapping 20 ms bins. This process resulted in a two-dimensional map containing the population correlation values for each time point and lag period. We created such maps separately for each recording session. To calculate the

population correlation maps for correct and error trials, we balanced the number of trials by setting a common value (50 trials) for each group. This process was repeated 25 times, with a different random subset of trials each time. The average of these 25 maps was taken as the final value for the respective session. Additionally, we calculated a null distribution for each session using the same procedure described above, except that the neuronal activity was shuffled for each area and time window. This process was repeated 500 times, and the resulting shuffled population correlation maps were used for a permutation test to determine whether the population correlation coefficient for each time bin could be explained by chance. The significance threshold was set to $p < 0.05$ (two-sided), meaning that the actual population correlation coefficient had to be greater than 97.5% of the values in the null distribution. We averaged the results across the 43 sessions that met our criteria and displayed the resulting population correlation map as a heatmap.

On the basis of population correlation maps, we computed for each session and each 20 ms time step a feedforward ratio in order to see at which time point the neuronal activity is more driven by a feedforward or feedback signal. The feedforward ratio is defined as follows:

$$Feedforward\ ratio = \frac{(AUC_{\text{feedforward}} - AUC_{\text{feedback}})}{(AUC_{\text{feedforward}} + AUC_{\text{feedback}})},$$ where $AUC_{feedforward}$ is the area under the positive-lag sides (or curve) of the population correlation map and $AUC_{feedback}$ is the area under the negative-lag sides of the population correlation map. By creating a shuffled distribution, we were able to see which ratio was significantly above chance or not. To do this, we eliminated the correspondence whether the calculated population correlation coefficients belong to the negative or positive lag. We performed this process 500 times and for each time bin separately. The significance threshold was set to $p < 0.05$ (two-sided). Furthermore, we compared the AUC values for the positive and negative lags between the correct and error trials by using a two-sided Wilcoxon signed-rank test. Analogous to what was described above, we used the shuffled distribution of population correlation maps to create a significance threshold ($p < 0.05$, two-sided).

Finally, we calculated mean population correlation functions which refer to the average of all population correlation functions across the time lags between the brain areas (−40 ms till +40 ms). These functions were constructed for correct and error trials und compared by a two-sided Wilcoxon signed-rank test. For visualization, the differences between the functions were plotted.

Centroid-based lag estimation

To characterize the temporal structure of inter-areal correlations, we quantified the lag at which canonical correlations peaked for each time bin during correct trials. For each bin, the lag of the maximum CCA value was identified from the time-lagged correlation matrix. To assess the robustness of these lag estimates, we computed a centroid around each peak, defined as the range of lag values encompassing the 95th percentile of CCA values centered on the maximum. The centroid provides a measure of the temporal spread of the correlation peak and was used for visualization. All analyses used the same preprocessing and CCA parameters as in the main analyses.

Population correlation analysis for specific neuronal populations

To examine the contribution of numerosity-selective neuronal populations to inter-areal communication, we repeated the canonical correlation analysis after stratifying multi-unit activity (MUA) by task selectivity. MUAs were classified as sample-selective, delay-selective, or non-selective based on numerosity tuning during the sample and/or delay period, using the same criteria as our single-unit analyses (one-way ANOVA, factor numerosity, $p < 0.01$; see Machts et al.[20]). Neuronal responses during the sample period were quantified in a 600-ms window, shifted 100 ms after sample onset to account for typical visual response latency. Delay-period activity was analyzed in a 1000-ms window, shifted 100 ms after sample offset.

All recording sessions included in the main correct/error trial analyses were also used here. However, stratifying by selectivity reduces the number of available MUAs per class. Therefore, the analysis required a minimum of one MUA per cortical area (VIP and dlPFC) and selectivity class per session, rather than the four-MUA minimum used in the main analyses.

MUAs selective during both sample and delay periods were included in both the sample- and delay-selective groups. Using this criterion, heatmaps of time-lagged CCA values were constructed from 43 sessions: 29 sessions with non-selective MUAs, 38 with sample-selective MUAs, and 40 with delay-selective MUAs. Apart from the adjusted inclusion criterion, the CCA procedure, preprocessing, and statistical analyses were identical to those in the main analyses.

Publication 3: Reaction-time signatures reveal divergent cognitive strategies underlying numerical decisions in monkeys and crows.

This chapter is identical to the published version of:

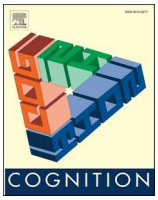
Machts, T., Kirschhock, M. E., & Nieder, A. (2026). Reaction-time signatures reveal divergent cognitive strategies underlying numerical decisions in monkeys and crows. Cognition, 273, 106527. https://doi.org/10.1016/j.cognition.2026.106527

Cognition 273 (2026) 106527

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit



Full Length Article

Reaction-time signatures reveal divergent cognitive strategies underlying numerical decisions in monkeys and crows

Tobias Machts [1], Maximilian E. Kirschhock [1,2], Andreas Nieder [*]

Animal Physiology, Institute of Neurobiology, University of Tuebingen, Auf der Morgenstelle 28, 72076 Tuebingen, Germany

ARTICLE INFO

Keywords:
Primate
Corvid
Numerosity
Decision making
Reaction-time dynamics

ABSTRACT

Non-symbolic numerical competence, supported by the approximate number system (ANS), is widespread across animals, yet it remains unclear whether similar numerical performance reflects shared cognitive mechanisms. We compared two rhesus monkeys and two carrion crows on an extended delayed match-to-numerosity task. Animals viewed a sample numerosity (one to four), maintained it across a delay, and searched sequentially through up to three test displays for a match. Behavioral accuracy was comparable across species, with both showing classical ANS signatures such as numerical distance and numerical size effects. However, reaction-time analyses revealed striking species differences: monkeys' responses slowed with increasing numerosity and later test phases, whereas crows responded faster over later test phases and showed numerosity-invariant reaction times. Classifier decoding further confirmed that numerosity and test phase were strongly reflected in monkeys' reaction times but only weakly in crows'. These findings demonstrate that equivalent numerical performance can arise from fundamentally different cognitive strategies, as revealed by distinct reaction-time dynamics, highlighting how evolutionary constraints shape species-specific solutions to the same computational problem.

One-sentence summary: Monkeys solve an abstract-category delayed match-to-sample task through slow, capacity-limited, sequential checking, whereas crows rely on fast, parallel, anticipatory processing with reduced working-memory load.

1. Introduction

Non-symbolic numerical competence is widespread across animals and supports adaptive decision-making (Brannon & Terrace, 1998; Howard, Avarguès-Weber, Garcia, Greentree, & Dyer, 2018; Nieder, 2020; Potrich, Zanon, & Vallortigara, 2022; Rugani, Fontanari, Simoni, Regolin, & Vallortigara, 2009; Scarf, Hayne, & Colombo, 2011; Stancher, Rugani, Regolin, & Vallortigara, 2015). Among vertebrates, both monkeys (primates) and crows (corvids) can discriminate, remember, and act upon numerical quantities (Cantlon & Brannon, 2006; Ditz & Nieder, 2016; Merten & Nieder, 2009). These behaviors are commonly attributed to the approximate number system (ANS), a domain-general mechanism that allows numerical estimation when explicit counting is not possible (Nieder, 2025, 2026). As a signature of the ANS, monkeys and corvids follow similar psychophysical principles — including the numerical distance effect (performance improves as the numerical difference between quantities increases), the numerical size effect (performance declines as overall magnitude increases) (Ditz & Nieder, 2016; Merten & Nieder, 2009), and systematic cognitive judgment biases (Jannasch, Grüb, & Nieder, 2025).

In addition to the ANS, numerical cognition research has proposed a second mechanism, the object tracking system (OTS), which is thought to support precise representation of small numerosities by assigning markers (or pointers) to small numbers of individual elements (typically up to ~3–4 items) (Pylyshyn, 2001). Whereas the ANS is characterized by ratio-dependent performance consistent with Weber's law, OTS processing is typically associated with rapid and accurate individuation of small sets without ratio dependence. Both systems are thought to coexist and may contribute differentially depending on task demands and numerical range (Feigenson, Dehaene, & Spelke, 2004).

So far, it remains unclear how closely the behavioral signatures of numerosity discriminations in monkeys and crows align under identical task conditions; whether the two species share not only the same performance outcomes but also the same cognitive strategies. Primates and

* Corresponding author.
E-mail address: andreas.nieder@uni-tuebingen.de (A. Nieder).
[1] Shared first authors.
[2] Present address: Department of Brain Sciences, Weizmann Institute of Science, Rehovot, Israel.

https://doi.org/10.1016/j.cognition.2026.106527
Received 19 January 2026; Received in revised form 9 March 2026; Accepted 11 March 2026
Available online 19 March 2026

songbirds differ in telencephalic brain neuroanatomy (layered cortex vs. nuclearly organized pallium, respectively) and sensorimotor systems (eye–hand vs. eye–beak). These differences might indicate that even when performing the same task, the species may rely on distinct cognitive strategies.

To address this question, we trained two rhesus monkeys and two carrion crows on an extended delayed match-to-numerosity task. Animals viewed a briefly presented sample numerosity, maintained it across a delay, and then searched for the matching numerosity in up to three sequential test displays. By varying target numerosity (one to four) and the number of nonmatching items, the task exhibited different demands on numerical discrimination, working memory, and recognition.

These manipulations are known to affect cognitive processes in humans, suggesting that similar cognitive constraints could influence nonhuman animals. For example, human studies show that multiple nonmatch distractors can induce sequential search, slowing reaction times and reducing accuracy (Pomè, Anobile, Cicchini, Scabia, & Burr, 2019; Railo, Koivisto, Revonsuo, & Hannula, 2008; Vetter, Butterworth, & Bahrami, 2008). Likewise, maintaining a visuo-spatial working-memory load while judging multi-item arrays reduces numerosity precision (Castaldi, Piazza, & Eger, 2021; Ester, Ho, Brown, & Serences, 2014). Together, these findings imply that task complexity interacts with memory and attention to shape numerosity judgments. This motivates a cross-species comparison to determine whether monkeys and crows are subject to similar cognitive constraints. Specifically, comparing their respective performances under identical task conditions can reveal whether similar accuracy reflects shared mechanisms or different processing styles, and how neural and memory constraints shape reaction-time patterns.

2. Results

2.1. Behavioral accuracy

Two monkeys and two crows were trained on a modified delayed match-to-numerosity task (Machts, Grüb, & Nieder, 2025; Machts & Nieder, 2026). They viewed sample numerosities from one to four, followed by a delay. Each trial then included up to three sequential test phases presented in equal proportion (see Materials and Methods, Fig. 1A). In each phase, the animals searched for a numerosity that matched the sample. They were rewarded for selecting a matching numerosity in any of the test displays, which encouraged them to inspect each stimulus carefully and continue searching until they found the match.

Both the monkeys and the crows mastered the task (Fig. 2). They reached similar mean performance levels across sessions, all well above

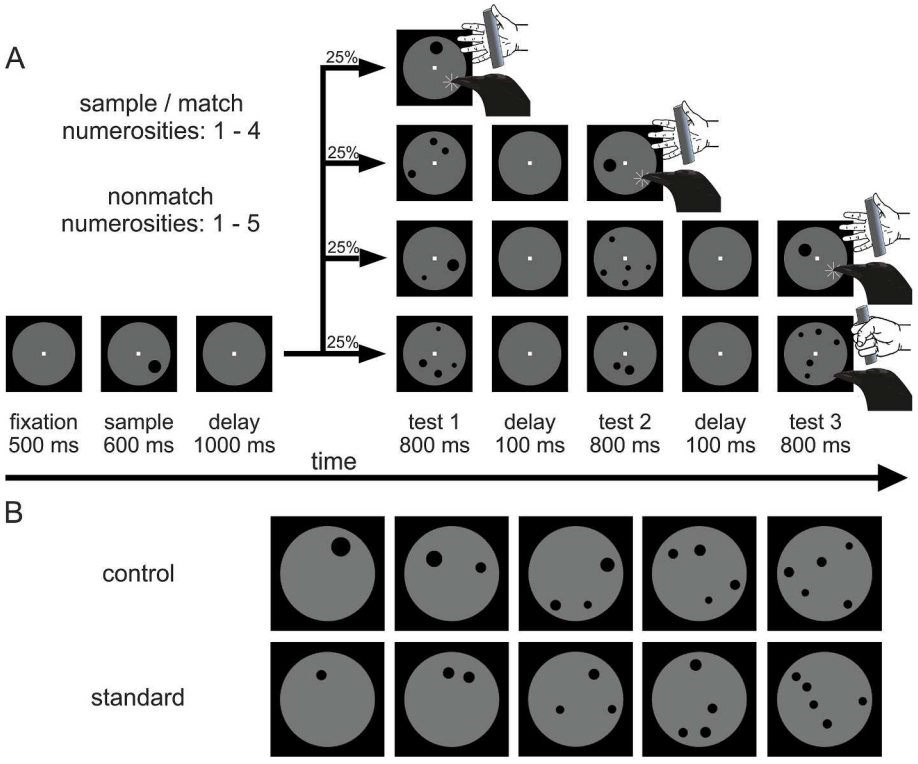


Fig. 1. Task protocol. (A) Both monkeys and crows performed a delayed match-to-sample task involving dot numerosities. The monkeys initiated a trial by grabbing the response bar and fixating on the fixation target, whereas the crows had to position their heads centrally in front of the monitor to close an infrared light barrier. These actions had to be maintained throughout the task, until the monkeys released the bar or the crows pecked at the touchscreen. During a sample phase, a sample stimulus with numerosities ranging from 1 to 4 was presented, followed by a delay phase. The monkeys and crows had to respond whenever a numerosity matching the sample was presented in one of three subsequent test phases. The monkeys had to respond by releasing the bar and the crows had to respond by moving the head out of the IR-light barrier. They had to refrain from responding whenever nonmatching numerosities (ranging from 1 to 5) were presented in the test phases. Up to three sequential test phases (Test 1 – Test 3) were presented pseudo-randomly in different trial conditions; in each test phase, a matching or nonmatching numerosity could appear. If the final Test 3 phase showed a nonmatch, the monkey had to maintain the bar and the crow had to withhold the pecking response until after the display offset. This resulted in a total of four different trial conditions. Sample numerosities and trial conditions were presented in a random and balanced way. (B) The spatial arrangement and diameter of the dots varied depending on the stimulus protocol. The standard stimuli featured randomly positioned and sized dots. In 50% of the trials, the standard stimuli were used, whereby an increasing number of dots resulted in an increasing occupied area. Therefore, control stimuli were presented in the remaining 50% of trials. In these, the dot density and total occupied area were equated across numerosities. Dot density was defined as the average distance between the centers of all dots on a numerosity display. Total occupied area was defined as the cumulative surface area of all the dots on a numerosity display. Thus, as the numerosity increased, the diameter of the dots decreased in the control stimuli. Both standard and control stimuli were presented pseudo-randomly and with equal frequency.

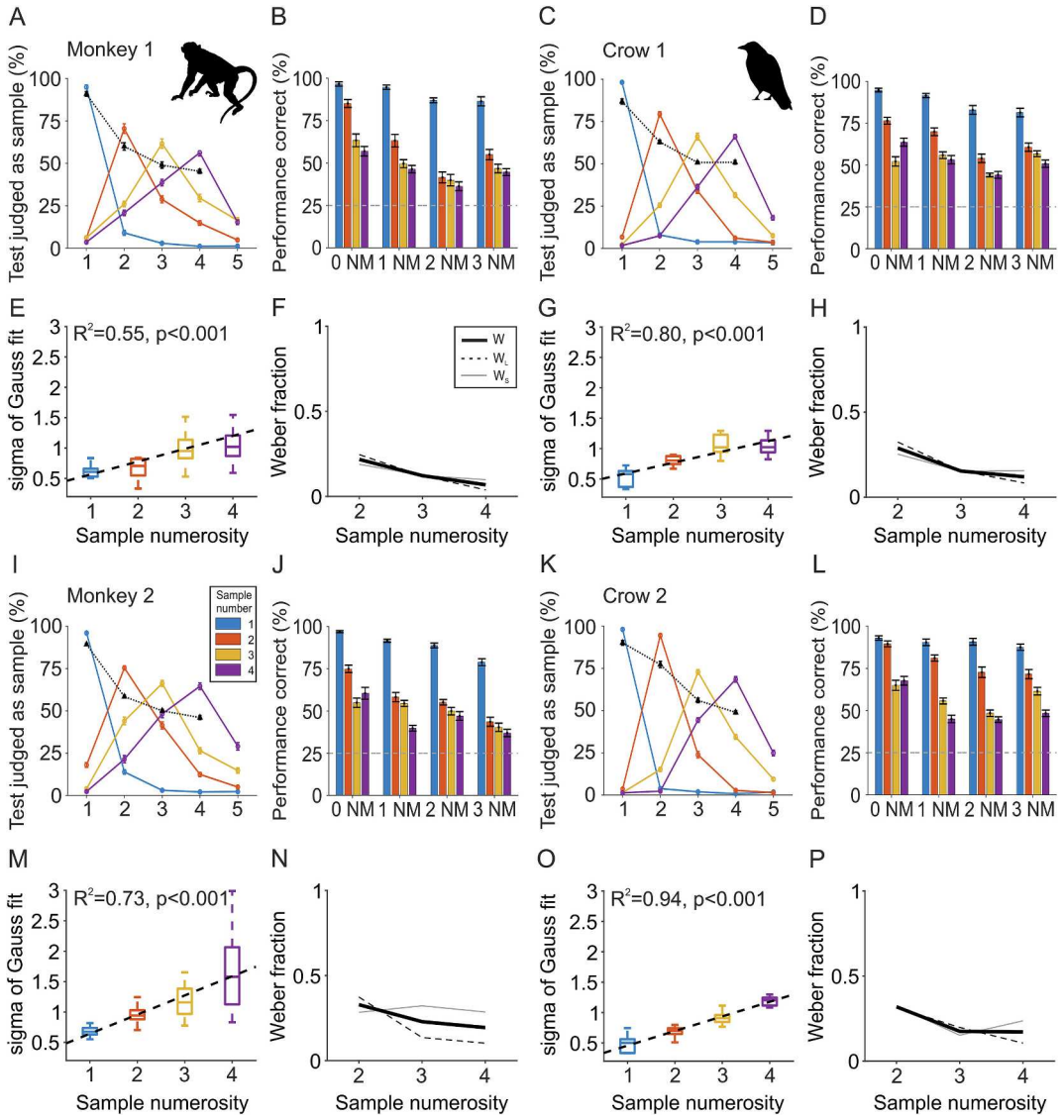


Fig. 2. Behavioral performance. (A, C, I, K) The behavioral performance functions of two monkeys and two crows are shown for sample numerosities of 1–4 (Monkey 1: 27 sessions; Monkey 2: 33 sessions; Crow 1: 20 sessions; Crow 2: 20 sessions). These functions reflect the probability that a monkey or a crow will judge a display in the test period as containing the same number of items as the sample numerosity (indicated by the solid line in four colors). The peak data points of each colored function indicate correct sample matching. Data points in the flanks reflect incorrect responses to non-matching numerosities. The black dotted line indicates the average performance for each sample numerosity. The error bars indicate the standard error of the mean (SEM) across sessions. (B, D, J, L) Average correct performance of the two monkeys and two crows to the four sample numerosities and the four test conditions (NM = nonmatch). Grey dashed line indicates chance level at 25%. Error bars indicate the SEM across sessions. (E, G, M, O) Distributions of widths (σ of fitted Gauss curves) for each sample numerosity 1 to 4 on linear numerical scaling are plotted as box plots (horizontal line indicate the median values). Dashed black line show the linear fit of Gauss width as a function of sample numerosity. (F, H, N, P).Mean Weber fractions (W, solid black line) for smaller (W_S, solid grey line) and larger ($_WL$, black dashed line) sample numerosity discrimination. Animal silhouettes available from phylopic.org under creative commons license.

the 25% chance level (Monkey 1: 27 sessions, 61.32 ± 1.36%, $p < 0.001$; Monkey 2: 33 sessions, 60.95 ± 0.58%, $p < 0.001$; Crow 1: 20 sessions, 59.90 ± 0.64%, $p < 0.001$; Crow 2: 20 sessions, 64.30 ± 1.14%, $p < 0.001$; Binomial test). Furthermore, to assess the animals' ability to identify the correct matching numerosity while ignoring nonmatching numerosities, we computed behavioral performance functions for each session. These functions quantify the probability that monkeys and crows judged a test numerosity as equal to the sample. Across all sample numerosities, both species selected the matching numerosity more often than any nonmatching numerosity (Fig. 2A, C, I, K).

Consistently, performance was highest for numerosity 1 (Monkey 1: 95.02 ± 1.12%; Monkey 2: 96.01 ± 0.52%; Crow 1: 98.12 ± 0.47%; Crow 2: 98.02 ± 0.48%). Accuracy declined as numerosity increased. For numerosity 4, peak performance dropped to 56.31 ± 1.96% (Monkey 1), 64.57 ± 1.99% (Monkey 2), 66.08 ± 1.37% (Crow 1), and 68.45 ± 1.63% (Crow 2). As predicted by the numerical distance effect — which posits that neighboring numerosities are the hardest to discriminate (Ditz & Nieder, 2016; Kirschhock, Ditz, & Nieder, 2021; Merten & Nieder, 2009; Nieder & Miller, 2003) — both monkeys and crows made the most errors when nonmatching numerosities were adjacent to the target. Their accuracy improved noticeably as the nonmatch numerosities became more distant from the target. Additionally, performance functions systematically widened as target numerosities increased, reflecting the numerical size effect, which indicates that discriminating larger numerosities is more difficult at a given numerical distance (Nieder, 2020). These patterns show that both species exhibit the hallmarks of the approximate number system (ANS) — a non-symbolic estimation system also observed in humans when counting is

prevented and they must rely on estimating numerosity (Merten & Nieder, 2009; Nieder, 2026).

Next, we calculated the average performance for each trial type—no nonmatch, one nonmatch, two nonmatches, and three nonmatches across all four sample numerosities (Fig. 2B, D, J, L). As expected, the number of nonmatch numerosities affected both species' ability to recall and identify the matching numerosity (three-way ANOVA with factors 'number of nonmatches', 'sample numerosity' and 'stimulus protocol'; Monkey 1: $F(3,841) = 79.74$, $p < 0.001$; Monkey 2: $F(3,1033) = 73.45$, $p < 0.001$; Crow 1: $F(3,617) = 70.63$, $p < 0.001$; Crow 2: $F(3,617) = 61.78$, $p < 0.001$). Compared to trials with no nonmatch numerosities, both monkeys' and crows' overall performance declined progressively as the number of nonmatches increased. This drop in performance likely reflects the growing cognitive demands on recognition memory (Machts et al., 2025). Monkey 1, Crow 1, and Crow 2—but not Monkey 2—showed a slight, idiosyncratic performance increase for numerosities following three nonmatches.

Furthermore, behavioral performance in both species depended on the sample numerosity within each trial type, shown by a significant interaction between trial type and sample numerosity (Monkey 1: $F(9,841) = 6.85$, $p < 0.001$; Monkey 2: $F(9,1033) = 4.71$, $p < 0.001$; Crow 1: $F(9,617) = 8.00$, $p < 0.001$; Crow 2: $F(9,617) = 10.63$, $p < 0.001$). Reflecting the numerical size effect, trials with numerosity 1 were the easiest for both monkeys and crows, while trials with numerosity 4 were the most challenging for the monkeys (Fig. 2B, D, J, L). Despite this, average performance across all trial types and sample numerosities was significantly above chance ($p < 0.05$, Binomial test), showing that both species were able to identify the correct numerosity most of the time.

Additionally, we used the three-way ANOVA to evaluate whether non-numerical visual cues influenced performance. The control protocol equated dot density and total occupied area across numerosities to minimize potential reliance on continuous visual variables. In crows, the analysis revealed no significant main effect of stimulus protocol and no significant interactions involving stimulus protocol (all $p > 0.05$), indicating that performance was indistinguishable between standard and control stimuli.

In monkeys, we observed statistically significant main effect of stimulus protocol (Monkey 1: $F(1,1033) = 35.51$, $p = 0.0353$; Monkey 2: $F(1,841) = 4,44$, $p < 0.001$). However, the magnitude of this effect was minute: performance differences between protocol types were only approximately 3–4% (Monkey 1: standard 61.3%, control 58.2%; Monkey 2: standard 63.3%, control 59.2%). Consistent with this, the corresponding effect sizes were small (Monkey 1: partial $\eta^2 < 0.01$; Monkey 2: partial $\eta^2 \approx 0.03$), indicating that the influence of stimulus protocol accounted for less than 1% in Monkey 1 and around 3% in Monkey 2 of the variance in behavioral performance.

In addition, weak interactions involving stimulus protocol were observed in individual animals. In one monkey, stimulus protocol interacted with the number of nonmatches (Monkey 1: $F(3,1033) = 5.21$, $p < 0.01$, partial $\eta^2 \approx 0.01$), whereas in the other monkey an interaction between stimulus protocol and sample numerosity was detected (Monkey 2: $F(3,841) = 7.18$, $p < 0.001$, partial $\eta^2 \approx 0.02$). However, these effects were small in magnitude, with effect sizes indicating that the interactions accounted for only a minor proportion of the variance. Overall, these results suggest that continuous visual variables such as dot density or total occupied area had only a limited influence on behavioral performance and cannot account for the main numerical effects observed in the task.

To formally quantify the numerical scaling properties of the behavioral performance functions, we fitted Gaussian functions to the tuning curves for each sample numerosity and estimated the width parameter (σ) of the fits. The width of the tuning curves increased systematically with increasing numerosity, indicating broader discrimination functions for larger numerosities (Fig. 2E,G,M,O). To test this relationship statistically, we correlated σ with sample numerosity. For all animals, σ increased significantly with numerosity (Pearson correlation: Monkey 1: $r = 0.55$, $p < 0.001$; Monkey 2: $r = 0.73$, $p < 0.001$; Crow 1: $r = 0.80$, $p < 0.001$; Crow 2: $r = 0.94$, $p < 0.001$).

In addition, we calculated Weber fractions derived from the behavioral tuning functions (Fig. 2F, H, N, P). These remained roughly constant across numerosities (~0.25), consistent with Weber-like scaling. Together, these results demonstrate that behavioral performance follows classical signatures of the approximate number system (ANS).

2.2. Reaction times

Although both species showed similar behavioral performance, reaction times (RTs) differed clearly between monkeys and crows. RT was measured for each matching numerosity and defined as the time from display onset to the animal's response — bar release for monkeys or touchscreen peck for crows (display duration: 800 ms). To account for visual processing and motor preparation, only responses of at least 100 ms after display offset were allowed (see Materials and Methods).

Both monkeys showed a positive relationship between RT and test phase for all four matching numerosities: they responded more slowly as the test phase increased (Fig. 3A, C). This effect was strongest for numerosities 2–4 and weaker or absent for numerosity 1 (Pearson

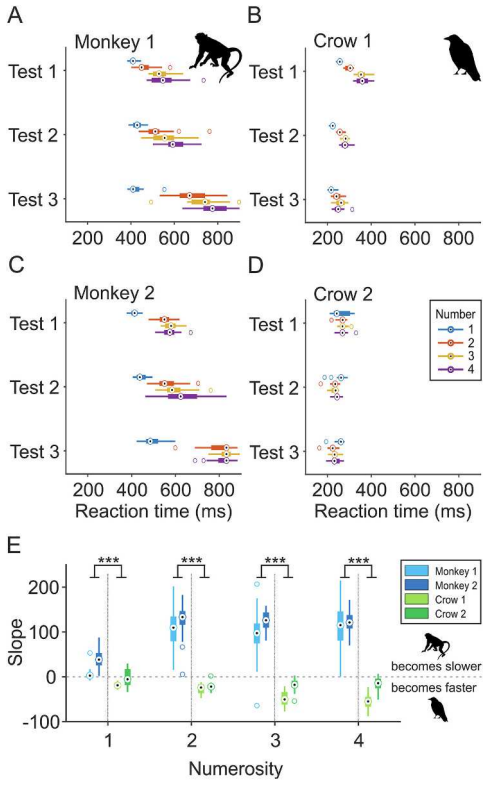


Fig. 3. Reaction times (RT) for correct responses to the matching test stimulus, categorized by the three test phases. (A, B, C, D) The colored boxplots indicate the RT for the respective matching numerosities (1–4) that the monkeys and crows should match to. The edges of the boxes represent the 25th and 75th percentiles, and the black dot inside the boxes demark the respective medians. Outliers are indicated by colored circles. Note that the monkeys become slower as the number of tests increases and the crows become faster (Monkey 1: 27 sessions; Monkey 2: 33 sessions; Crow 1: 20 sessions; Crow 2: 20 sessions). (E) The RTs for the three test phases were fitted using a linear function (linear regression). The resulting slopes were plotted as boxplots for the respective matching numerosities (1–4) and compared between monkeys and crows using a Wilcoxon rank sum test. A positive slope indicates that the animals are getting slower with increasing test phase, while a negative slope indicates that they are getting faster. See (A, B, C, D) for boxplot description. ***: $p < 0.001$. Animal silhouettes available from phylopic.org under creative commons license.

correlation; Monkey 1: numerosity 1: $r = 0.17$, $p = 0.136$; numerosity 2: $r = 0.79$, $p < 0.001$; numerosity 3: $r = 0.73$, $p < 0.001$; numerosity 4: r = 0.79, $p < 0.001$; Monkey 2: numerosity 1: r = 0.73, p < 0.001; numerosity 2: $r = 0.81$, p < 0.001; numerosity 3: $r = 0.84$, p < 0.001; numerosity 4: $r = 0.82$, p < 0.001).

Interestingly, the pattern was reversed in crows (Fig. 3B, D). Both crows showed a negative relationship between RT and test phase: they responded faster as the test phase increased. Pearson correlations were as follows: Crow 1: numerosity 1: $r = -0.78$, p < 0.001; numerosity 2: $r = -0.80$, p < 0.001; numerosity 3: $r = -0.86$, p < 0.001; numerosity 4: r = −0.86, p < 0.001; Crow 2: numerosity 1: $r = -0.07$, $p = 0.608$; numerosity 2: $r = -0.68$, p < 0.001; numerosity 3: $r = -0.63$, p < 0.001; numerosity 4: $r = -0.51$, $p < 0.001$.

This pattern was confirmed by fitting a linear function to the RTs for each matching numerosity across the three test phases and calculating the slopes (Fig. 3E). Within each species, slope steepness was qualitatively similar across numerosities. Monkeys showed positive slopes, indicating slower responses with later test phases, whereas crows showed negative slopes, indicating faster responses. The slopes differed highly and significantly between monkeys and crows (numerosity 1: Z = −6.564, p < 0.001; numerosity 2: Z = −8.411, p < 0.001; numerosity 3: Z = −8.154, p < 0.001; numerosity 4: Z = −8.396, p < 0.001; Wilcoxon rank sum test, two-sided). Notably, slope steepness was much weaker for numerosity 1 compared to the other numerosities in both species — especially in monkeys (Monkey 1: 3.17; Monkey 2: 38.69; Crow 1: −19.00; Crow 2: −5.50) — indicating that responses to numerosity 1 were less influenced by increasing test phases.

We also observed differences in how the animals responded to different numerosities within a test phase. Monkeys were fastest for numerosity 1 and progressively slower for higher numerosities (Fig. 3A, C). Monkey 1's median RTs for numerosity 1 were 409.63 ms, 428.70 ms, and 409.83 ms across the three test phases; Monkey 2's were 413.88 ms, 437.51 ms, and 485.32 ms. In contrast, both monkeys responded slowest to numerosity 4. Monkey 1's RTs were 547.23 ms, 592.78 ms, and 775.68 ms; Monkey 2's RTs were 574.59 ms, 624.23 ms, and 833.19 ms.

Notably, this gradation was absent in both crows, which responded to all four numerosities with similar RTs (Fig. 3B, D). For Crow 1, median RTs for numerosity 1 were 256.75 ms, 224.75 ms, and 217.25 ms across the three test phases, and for numerosity 4 they were 242.75 ms, 262.25 ms, and 261.75 ms. For Crow 2, median RTs for numerosity 1 were 359.50 ms, 280.75 ms, and 249.50 ms, and for numerosity 4 they were 269.00 ms, 244.00 ms, and 232.00 ms.

2.3. SVM classifier predicts numerosity and test phase based on the reaction times

Collectively, our analyses indicate that, unlike carrion crows, rhesus monkeys respond distinctively to both numerosity and test phase in our task. To confirm this, we used an SVM classifier to predict numerosity and test phase based on each animal's reaction times, training the classifier separately for each subject with 7-fold cross-validation and 150 resampling iterations. Because decoding accuracy depends on sample size, we created pseudo-populations to balance the number of sessions between monkeys and crows. The SVM classifier performed well above the 8.33% chance level for both monkeys and crows (Fig. 4A). However, monkeys achieved significantly higher decoding accuracy than crows (Z = −19.0613, $p < 0.001$, Wilcoxon rank sum test, two-sided). Decoding success was similar between the two monkeys, with no significant difference (Monkey 1: 43.94%, Monkey 2: 44.84%; Z = −0.5394, $p = 0.590$). In crows, decoding success was higher in Crow 1 (32.68%) than in Crow 2 (27.10%; Z = 8.3495, p < 0.001, Wilcoxon rank sum test, two-sided).

In addition to overall decoding accuracy, we generated confusion matrices for each animal from the classifier's predictions. In monkeys, the highest accuracy clearly aligned along the main diagonal (Fig. 4B, D), indicating reliable classification. This pattern was absent in crows,

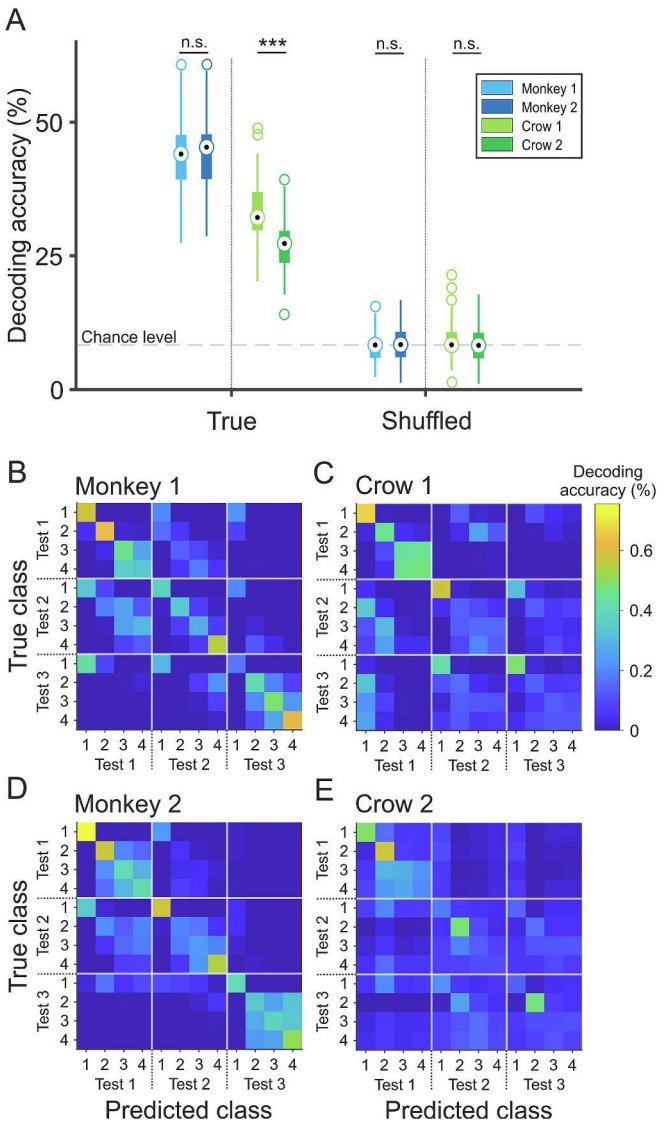


Fig. 4. Decoding numerosity and test phase from RT with a SVM classifier. (A) Mean decoding accuracy of the classifier for true and shuffled labels. The edges of the boxes represent the 25th and 75th percentiles, and the black dot inside the boxes the respective medians. Outliers are indicated by colored circles. Grey dashed line indicates chance level at 8.33%. Decoding accuracy was tested between the animals using a Wilcoxon rank sum test. ***: p < 0.001; n.s.: not significant. (B-E) Confusion matrices for sessions with at least 7 RTs for each correct match numerosity and test phase (Monkey 1: 18 sessions; Monkey 2: 29 sessions; Crow 1: 20 sessions; Crow 2: 20 sessions). Rows within the confusion matrices correspond to classification performance curves per numerosity and test phase, which was evaluated by a seven-fold cross-validation within the sessions. The sessions count was balanced across the animals by creating pseudo populations. The scaling of color maps is the same across the four confusion matrices. Color values in the confusion matrices show the average over cross-validation runs and 150 resamples. Note that for both monkeys, the classifier can predict the correct match numerosity of the respective test phase based on the RT, whereas this is not possible for neither crows.

where decoding of individual classes was indistinguishable (Fig. 4C, E). These results show that an ideal observer could predict numerosity and test phase from RTs far more accurately in rhesus monkeys than in carrion crows.

2.4. Accuracy in relation to reaction times

Finally, we examined the relationship between accuracy and

reaction time using Pearson correlations for each numerosity and test phase. Both monkeys tended to respond faster when performing well, although this pattern was not consistent across all numerosities and test phases and was more pronounced in Monkey 1 (Fig. 5A, C). No significant cases were found where monkeys were slower when performing well. In contrast, crows tended to be slower when performing well, with this effect being more noticeable in Crow 2 (Fig. 5B, D). As with monkeys, this opposite effect — faster responses with better performance — was not statistically significant in crows. Importantly, faster RTs in crows were not associated with reduced accuracy; if anything, higher performance tended to co-occur with shorter reaction times, arguing against a simple speed–accuracy tradeoff.

3. Discussion

We found that monkeys' reaction times (RTs) increased with later test phases and larger numerosities, whereas crows' RTs decreased with test phase and remained largely invariant across numerosities. SVM decoding revealed strong structure in monkeys' RTs but only weak structure in crows', while overall accuracy was comparable across species. These results indicate that, despite similar performance levels, monkeys and crows solve the same numerical task using fundamentally different cognitive strategies. Both species show signatures of the approximate number system (ANS), yet their opposing temporal profiles suggest that monkeys' behavior is best explained by slow, memory-dependent, sequential evaluation, whereas crows' behavior is consistent with fast, parallel processing strategies.

Because the present study included numerosities up to five, it is conceivable that object tracking system (OTS)-related mechanisms contributed to performance. However, the behavioral signatures observed here even for small numerosities—most notably the ratio-dependent discrimination performance consistent with Weber's law—align more closely with predictions of ANS-based processing. This is consistent with previous studies reporting ratio-dependent discrimination across small and large numerosities in these trained species (Cantlon & Brannon, 2006; Ditz & Nieder, 2015, 2016; Jannasch et al., 2025; Jordan & Brannon, 2006; Liao, Brecht, Veit, & Nieder, 2024; Merten & Nieder, 2009; Nieder & Miller, 2003). Given this shared Weber-like scaling across numerosities, we would therefore expect the species-specific reaction time dynamics observed here to persist even for larger sample numerosities.

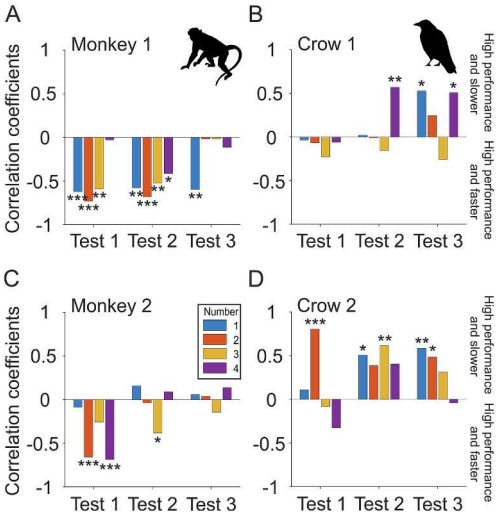


Fig. 5. Pearson correlation coefficients showing the relationship between performance and RT. (A, B, C, D) A Pearson correlation was used to investigate the relationship between performance and RT of monkeys and crows. The resulting correlation coefficients (Pearson's r) were plotted for each correct match numerosity and test phase. Both monkeys tend to have a negative correlation, i.e. they have a better performance when they are faster. In contrast, crows tend to have a positive correlation, i.e. they have a better performance, when they are slower. Stars above or below the bars indicate statistical significance for the respective correlation coefficient. *: $p < 0.05$, **: $p < 0.01$, ***: $p < 0.001$. Animal silhouettes available from phylopic.org under creative commons license.

3.1. Monkeys use a slow, sequential, memory-limited search strategy

Monkeys responded more slowly when the target appeared later in the sequence and when numerosity increased. The SVM classifier could reliably decode both numerosity and test phase from monkeys' RTs, confirming that their behavior carries strong signatures of serial, memory-dependent evaluation that resembles controlled processing in humans. Their performance slows when more items must be processed or when the target appears later in a sequence, indicating that working-memory load increases over time (Yakovlev, Bernacchia, Orlov, Hochstein, & Amit, 2005). Monkeys also rely on serial, position-based encoding, as shown by ordinal-position errors in sequence recall tasks (Orlov, Yakovlev, Amit, Hochstein, & Zohary, 2002). Even in numerosity judgments — often thought to be fast and parallel — reaction times rise with numerical magnitude, suggesting that decisions become slower as quantity increases (Nieder & Miller, 2004). Monkeys also fail to extract relational or structural patterns from sequences and instead depend on item-specific working-memory representations, which require more effort and are less efficient than human strategies (Zhang et al., 2022). Neurophysiological studies further show that monkeys hold a target numerosity in working memory and check each test display in sequence, using sustained prefrontal and parietal activity to guide recognition (Machts et al., 2025). Together, these findings indicate that monkeys rely on slow, effortful, and capacity-limited processing when they must maintain information and evaluate items one by one. This aligns with evidence that primate numerical cognition depends on fronto-parietal working memory in sequential numerical tasks (Nieder, 2012; Nieder, Diester, & Tudusciuc, 2006).

We acknowledge that the motor responses required from monkeys and crows were not identical: monkeys released a bar, whereas crows moved their heads out of a light barrier. However, in both cases the task required an active, self-initiated movement to report the decision. We therefore consider it unlikely that these relatively small differences in response modality account for the relative differences in reaction times observed between the two species.

3.2. Crows appear to use a parallel, rapid, or anticipatory strategy

Crows, in contrast, responded faster with later test phases and showed near-constant RTs across numerosities, indicating that neither working-memory load nor numerical magnitude imposed measurable costs. Importantly, the speeding observed in crows across later test phases cannot be explained by a simple speed–accuracy tradeoff, as faster responses were not associated with reduced performance; if anything, higher accuracy tended to co-occur with shorter reaction times. This stands in sharp contrast to the monkeys' behavior. Crows also showed shallower SVM decoding accuracy: numerosity and test phase could still be decoded above chance, but with markedly weaker structure, consistent with less systematic variation in RTs.

This pattern matches a body of evidence suggesting that avian numerical processing is rapid, parallel, and feedforward-like. Even numerically naïve crows possess neurons tuned to visual numerosity (Wagener, Loconsole, Ditz, & Nieder, 2018), and young chicks show similar innate number-selective coding (Kobylkov, Mayer, Zanon, & Vallortigara, 2022). Crows also rely on approximate magnitude representations rather than serial counting across sequential and simultaneous inputs (Ditz & Nieder, 2020). Their numerosity-to-action behavior, performed with very short response latencies (≈389–444 ms), likewise suggests rapid stimulus–response mapping without heavy

memory involvement (Kirschhock & Nieder, 2022). These findings support the interpretation that crows rely on fast, low-cost perceptual strategies, with reduced dependence on sequential checking or sustained working memory (Bobrowicz & Osvath, 2019).

3.3. Different cognitive strategies despite similar accuracy

The opposing temporal signatures — load-dependent slowing in monkeys versus anticipatory speeding in crows — demonstrate that similar behavioral performance does not imply shared cognitive mechanisms. Monkeys appear to rely on capacity-limited, controlled processes, whereas crows employ fast, parallel, and heuristic strategies. This divergence reflects a broader comparative pattern: evolution can produce convergent cognitive abilities while relying on distinct neural architectures and computational solutions (Nieder, 2021; Olkowicz et al., 2016). Future work should examine whether these species differences persist under controlled temporal conditions, in the presence of distractors, or with increased working-memory demands.

4. Methods

4.1. Experimental animals

We trained two male adult rhesus monkeys (*Macaca mulatta*, named Monkey 1 and Monkey 2) and two adult carrion crows (*Corvus corone*, named Crow 1 and Crow 2) at the institute's facility and acquired data from them. The monkeys were housed socially under a 12/12-h light–dark cycle. They were provided with environmental enrichment for climbing and foraging, and had access to primate food at all times, with fresh fruit and vegetables provided at regular intervals. During the experiments, the monkeys were kept on a controlled water protocol and earned water as a reward during training.

The crows were housed in small social groups in large indoor aviaries. They were kept on a controlled feeding protocol and earned food as a reward during training. If necessary, water/food was supplemented after the sessions. All experimental procedures had been ethically reviewed and approved and were regulated and supervised by the national authorities, the Regierungspräsidium Tübingen, Germany.

4.2. Experimental apparatus and stimuli

Training sessions were conducted for both animal species in a darkened operant conditioning chamber. The monkeys were seated in primate chairs 57 cm from a 15-in. flat-screen monitor with a resolution of 1024 × 768 pixels and a refresh rate of 60 Hz. An infrared eye-tracking system (ISCAN, Cambridge, MA) was used to control eye fixation while presenting test stimuli. An automated water dispenser was used for reward delivery.

The crows were loosely strapped to a wooden perch with leather jesses for training and positioned in front of a 15-in. touchscreen monitor (3 M Microtouch; 60 Hz refresh rate). An infra-red (IR) light barrier ensured that the crows kept their heads still and maintained a central head position in front of the monitor at a viewing distance of 14 cm throughout the trial. This light barrier consisted of an infrared light emitter/detector fixed to the ceiling of the chamber and a reflector foil attached to the crow's head. The reward (Birdseed pellets and mealworms) was delivered by an automated feeder that was positioned below the monitor. Loudspeakers (Visation WB10) were installed in the chamber to provide auditory feedback, and an infrared camera (Genius iSlim 321R) was installed to enable observational control. The NIMH Cortex software (Bethesda, MD) was used to present the stimuli and acquire behavioral data in both set-ups.

We used random dot arrays as stimuli representing numerosities ranging from one to five dots. The diameter of these dots ranged from 0.5° to 0.9° for the monkeys and from 1.2° to 5.5° for the crows. Each unique dot pattern was presented on a grey background circle subtending 5.4° (for the monkeys) and 26.1° (for the crows) of visual angle. This pattern was used as both the sample stimulus and the test stimulus. To prevent the monkeys and crows from memorizing stimulus sequences, a MATLAB (The MathWorks, Natick, MA) program generated the stimuli anew before each training session. The sample and test stimuli of the same countable numerosity were never identical within a trial or session. To control for low-level, non-numerical visual features, dot diameters were randomly selected in half of the trials (standard protocol, Fig. 1B). In the other half of the trials, the dot density and the total area occupied were equated across stimuli (control protocol, Fig. 1B). Trials of the standard and control protocols were presented in a pseudo-random and unpredictable order.

4.3. Behavioral protocol

We trained two macaque monkeys and two carrion crows on a modified, extended version of the delayed match-to-numerosity task, in which the animals had to match the numerosity of the test stimulus to that of the sample stimulus (Fig. 1A). Nonmatching numerosities presented between the sample and the matching test stimulus had to be ignored. Up to three sequential numerical nonmatches were presented, i.e., the task consisted of up to three test phases, in each of which the correct or incorrect numerosity could be presented.

While the monkeys initiated a trial by grabbing a bar and maintaining eye fixation within a visual angle of 3.5° of a central white dot, the crows had to position their heads centrally in front of the monitor to close an infrared light barrier. Trials were immediately aborted and excluded from further analysis if the monkeys broke fixation or the crows moved out of the predefined position. Initially, a grey circular background was displayed for 500 ms (the pure fixation phase). Subsequently, the sample stimulus containing one to four dots appeared for 600 ms. This sample stimulus had to be memorized over a delay phase of 1000 ms (again with a grey circular background) and then compared in a subsequent sequence of test numerosities. Each test stimulus was presented for 800 ms, with a 100 ms delay between each stimulus.

If the numerosity of the test stimulus was the same as that of the sample stimulus (match situation: 1–4), the monkeys had to respond by releasing the bar and the crows had to respond by moving the head out of the IR-light barrier. If the numerosities were different (nonmatch situation: 1–5), the monkeys continued holding the bar and the crows maintained the predefined position.

A nonmatch numerosity was never shown twice per trial. Sample numerosities and trial types (i.e. whether no, one, two or three nonmatch numerosities were presented) were balanced across conditions and displayed with equal probability (25% each, pseudo-randomly intermixed). Thus, the animals could not predict whether or when the match stimulus would appear. A correct decision per trial was rewarded with a drop of water for the monkeys and a small piece of food for the crows. In 25% of trials involving three nonmatch numerosities in sequence, the monkeys were rewarded for holding the bar and the crows for maintaining the predefined position throughout the trial until the display offset. Since the number of trials that a monkey or crow can perform during an experimental session is limited (on average, Monkey 1 completed 483 trials, Monkey 2 completed 534 trials, Crow 1 completed 640 trials and Crow 2 completed 631 trials), not all possible combinations of sample (1–4) and test numerosities (1–5) could be presented. Therefore, a predefined collection of 64 unique sample-test combinations was presented instead.

Behavioral chance level calculation was as follows: For each test phase, the animals have only two options: to respond or to no respond, corresponding to a probability of $p = 0.5$. Cascading this probability for the subsequent test phases (by multiplication), the following probabilities arise: Test 1 phase: $p = 0.5$; Test 2 phase: $p = 0.25$; Test 3 phase: $p = 0.125$. Since the animals do not know how many test phases to expect (i.e., trial types are unpredictable), the average probability results in $p = 0.25$.

4.4. Data analysis

All analyses were conducted in MATLAB (version R2021a, MathWorks Inc.). All values in the main text and figures refer to the mean ± SEM. SEM was calculated as the standard deviation divided by the square root of the number of samples.

4.5. Behavioral performance

Behavioral performance (% correct) was calculated by counting the number of correctly performed trials in each session and dividing this by the total number of trials. Percent correct performance for each stimulus combination was calculated for each session to construct behavioral performance functions in response to each sample numerosity. Mean response functions were computed as the average of all individual functions per session.

To assess the effects of task demands on behavioral accuracy, we conducted three-way repeated-measures analyses of variance (ANOVA; Interaction Model) separately for each animal. The dependent variable was percent correct performance per session. The within-session factors were 'number of nonmatches' (0–3), 'sample numerosity' (1–4) and 'stimulus protocol' (standard, control). Furthermore, we computed their interaction term. Because we test distributions of means (% accuracy), we can assume normal distributions suitably tested with parametric tests. For all ANOVAs, we report F-values, degrees of freedom, and associated p-values. Effect sizes (partial η^2) were calculated to quantify the magnitude of observed effects.

To quantify the performance functions, we fitted Gaussian bell curves to the performance functions on a linear numerical scale. The center of each fitted Gaussian curve was fixed to the respective target number and the width, height and offset parameters were fitted to minimize the overall residual error.

We calculated Weber fractions to assess the just noticeable difference between two stimuli (Weber, 1851). The following formulas were used to calculate the smaller and larger slopes of the performance functions, respectively:

$W_S = \frac{n - n_S}{n_S}$ for smaller Weber fractions, and

$W_L = \frac{n_L - n}{n}$ for larger Weber fractions.

With n being the target number, n_S is a number smaller than the target number, which the monkeys and crows could discriminate at a rate of 50%. Likewise, n_L is a number greater than the target number and is discriminated in 50% of cases.

4.6. Reaction times (RTs)

RTs were measured for each match numerosity (1–4 in Test 1 – Test 3) and was defined as the time taken by the monkey or crow to react (by releasing the bar in monkeys or by moving the head out of the IR-light barrier in crows) to the matching numerosity after display onset (the display duration was 800 ms). Due to the visual response latency and the time required for motor preparation, the animals were permitted to respond to the stimulus for up to 100 ms after display offset. Consequently, using this approach, responses occurring 100 ms after onset were also counted as errors and excluded from further analysis. Thus, the crows and monkeys could react within a time window of 800 ms, 100 ms after stimulus onset.

We performed Pearson correlation to investigate the relationship between RT and increasing test phase for each match numerosity. Furthermore, we fitted linear functions to the RTs using the following formula:

$$f(x) = mx + n,$$

where m is the slope, n is the $f(x)$-intercept and x is the respective test phase. This was repeated for each session to obtain a distribution of slopes for each animal. These distributions were then compared across species using a Wilcoxon rank sum test (two-sided).

Finally, we conducted Pearson correlations to investigate the relationship between RT and behavioral performance as a percentage for each match numerosity and test phase.

4.7. SVM classifier

We trained a linear multiclass support vector machine (SVM) classifier using one-versus-one classification (Chang & Lin, 2011) on response times evaluated for four matching numerosities and three test phases (i.e. 12 different classes). We only used sessions in which we could determine at least seven RTs for each class. This criterion was met in 18 sessions with Monkey 1, 29 sessions with Monkey 2, and 20 sessions with Crow 1 and Crow 2. Seven-fold cross-validation was performed, resulting in six trials for training and one trial for testing per class. Within each cross-validation loop, a new training and test subset was selected. We repeated the training and testing process 150 times, selecting a new set of trials at random with each iteration. The tested SVM classifier provides predicted class labels for the trials in the test subset. This enabled us to compute the main decoding success, as well as construct confusion matrices in combination with the true class labels of the test subset. Both possess an accuracy measure expressed as a percentage. We evaluated the decoding performance of the SVM classifier in balanced pseudo-populations of sessions, drawing a new set of equal size with each iteration (the decisive factor was the session number of Monkey 1). To create a chance-level performance distribution, we randomly shuffled the association between firing rates and trial class labels (numerosity) with each iteration. We tested the distributions of decoding success between animals using a Wilcoxon rank sum test (two-sided, with alpha correction).

CRediT authorship contribution statement

Tobias Machts: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Investigation, Formal analysis, Data curation, Conceptualization. **Maximilian E. Kirschhock:** Writing – review & editing, Investigation, Conceptualization. **Andreas Nieder:** Writing – review & editing, Writing – original draft, Validation, Supervision, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

None.

Acknowledgments

This work was supported by a DFG grants NI 618/15-1 and NI 618/17-1 (FOR 5159) to A.N.

Data availability

The data and code underlying all the findings of this manuscript are available on Figshare (DOI: https://doi.org/10.6084/m9.figshare.31045399).

References

Bobrowicz, K., & Osvath, M. (2019). Cognition in the fast lane: Ravens' gazes are half as short as humans' when choosing objects. *Animal Behavior and Cognition, 6*(2), 81–97.

Brannon, E. M., & Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science, 282*(5389), 746–749.

Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science, 17*(5), 401–406.

Castaldi, E., Piazza, M., & Eger, E. (2021). Resources underlying visuo-spatial working memory enable veridical large numerosity perception. *Frontiers in Human Neuroscience, 15*, Article 751098.

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology, 2*, 27.

Ditz, H. M., & Nieder, A. (2015). Neurons selective to the number of visual items in the corvid songbird endbrain. *Proceedings of the National Academy of Sciences of the United States of America, 112*(25), 7827–7832.
Ditz, H. M., & Nieder, A. (2016). Numerosity representations in crows obey the Weber–Fechner law. *Proceedings of the Royal Society B: Biological Sciences, 283*(1827), 20160083.
Ditz, H. M., & Nieder, A. (2020). Format-dependent and format-independent representation of sequential and simultaneous numerosity in the crow endbrain. *Nature Communications, 11*(1), 686.
Ester, E. F., Ho, T. C., Brown, S. D., & Serences, J. T. (2014). Variability in visual working memory ability limits the efficiency of perceptual decision making. *Journal of Vision, 14*(4), 2.
Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences, 8*(7), 307–314.
Howard, S. R., Avarguès-Weber, A., Garcia, J. E., Greentree, A. D., & Dyer, A. G. (2018). Numerical ordering of zero in honey bees. *Science, 360*(6393), 1124–1126.
Jannasch, L. L., Grüb, J., & Nieder, A. (2025). Shared cognitive biases influence numerical judgments in macaques and crows. *Proceedings of the National Academy of Sciences of the United States of America, 122*(44), Article e2512219122.
Jordan, K. E., & Brannon, E. M. (2006). Weber's Law influences numerical representations in rhesus macaques (Macaca mulatta). *Animal Cognition, 9*(3), 159–172.
Kirschhock, M. E., Ditz, H. M., & Nieder, A. (2021). Behavioral and neuronal representation of numerosity zero in the crow. *Journal of Neuroscience, 41*(22), 4889–4896.
Kirschhock, M. E., & Nieder, A. (2022). Number selective sensorimotor neurons in the crow translate perceived numerosity into number of actions. *Nature Communications, 13*(1), 6913.
Kobylkov, D., Mayer, U., Zanon, M., & Vallortigara, G. (2022). Number neurons in the nidopallium of young domestic chicks. *Proceedings of the National Academy of Sciences of the United States of America, 119*(32), Article e2201039119.
Liao, D. A., Brecht, K. F., Veit, L., & Nieder, A. (2024). Crows "count" the number of self-generated vocalizations. *Science, 384*(6698), 874–877.
Machts, T., Grüb, J., & Nieder, A. (2025). Neuronal encoding of recognition memory for numerical quantities in macaque intraparietal and prefrontal cortices. *Progress in Neurobiology, 251*, Article 102794.
Machts, T., & Nieder, A. (2026). Coordinated parieto–frontal neuronal communication is critical for abstract quantity judgments in primates. *Cell Reports*. https://doi.org/10.1016/j.celrep.2026.117147. Advance online publication.
Merten, K., & Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. *Journal of Cognitive Neuroscience, 21*(2), 333–346.
Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences of the United States of America, 109*(29), 11860–11865.
Nieder, A. (2020). The adaptive value of numerical competence. *Trends in Ecology & Evolution, 35*(7), 605–617.
Nieder, A. (2021). The evolutionary history of brains for numbers. *Trends in Cognitive Sciences, 25*(7), 608–621.
Nieder, A. (2025). The calculating brain. *Physiological Reviews, 105*(1), 267–314.
Nieder, A. (2026). The making of number: From content to representation. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2025.12.011. S1364-6613(25)00360–2. Advance online publication.
Nieder, A., Diester, I., & Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Science, 313*(5792), 1431–1435.
Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron, 37*(1), 149–157.
Nieder, A., & Miller, E. K. (2004). Analog numerical representations in rhesus monkeys: Evidence for parallel processing. *Journal of Cognitive Neuroscience, 16*(5), 889–901.
Olkowicz, S., Kocourek, M., Lučan, R. K., Porteš, M., Fitch, W. T., Herculano-Houzel, S., & Němec, P. (2016). Birds have primate-like numbers of neurons in the forebrain. *Proceedings of the National Academy of Sciences of the United States of America, 113*(26), 7255–7260.
Orlov, T., Yakovlev, V., Amit, D., Hochstein, S., & Zohary, E. (2002). Serial memory strategies in macaque monkeys: Behavioral and theoretical aspects. *Cerebral Cortex, 12*(3), 306–317.
Pomè, A., Anobile, G., Cicchini, G. M., Scabia, A., & Burr, D. C. (2019). Higher attentional costs for numerosity estimation at high densities. *Attention, Perception & Psychophysics, 81*(8), 2604–2611.
Potrich, D., Zanon, M., & Vallortigara, G. (2022). Archerfish number discrimination. *eLife, 11*, Article e74057.
Pylyshyn, Z. W. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition, 80*, 127–158.
Railo, H., Koivisto, M., Revonsuo, A., & Hannula, M. M. (2008). The role of attention in subitizing. *Cognition, 107*(1), 82–104.
Rugani, R., Fontanari, L., Simoni, E., Regolin, L., & Vallortigara, G. (2009). Arithmetic in newborn chicks. *Proceedings. Biological sciences, 276*(1666), 2451–2460.
Scarf, D., Hayne, H., & Colombo, M. (2011). Pigeons on par with primates in numerical competence. *Science, 334*(6063), 1664.
Stancher, G., Rugani, R., Regolin, L., & Vallortigara, G. (2015). Numerical discrimination by frogs (Bombina orientalis). *Animal Cognition, 18*(1), 219–229.
Vetter, P., Butterworth, B., & Bahrami, B. (2008). Modulating attentional load affects numerosity estimation: Evidence against a pre-attentive subitizing mechanism. *PLoS One, 3*(9), Article e3269.
Wagener, L., Loconsole, M., Ditz, H. M., & Nieder, A. (2018). Neurons in the endbrain of numerically naive crows spontaneously encode visual numerosity. *Current Biology, 28*(7), 1090–1094.e4.
Weber, E. H. (1851). *Die Lehre vom Tastsinne und Gemeingefühle: Auf Versuche gegründet*. Braunschweig, Germany: Friedrich Vieweg und Sohn.
Yakovlev, V., Bernacchia, A., Orlov, T., Hochstein, S., & Amit, D. (2005). Multi-item working memory–a behavioral study. *Cerebral Cortex, 15*(5), 602–615.
Zhang, H., Zhen, Y., Yu, S., Long, T., Zhang, B., Jiang, X., … Wang, L. (2022). Working memory for spatial sequences: Developmental and evolutionary factors in encoding ordinal and relational structures. *Journal of Neuroscience, 42*(5), 850–864.