Rethinking Inner Speech through Linguistic Active Inference

Bo Yao[1*]

1 Department of Psychology, Faculty of Science and Technology,

Lancaster University, Bailrigg, Lancaster LA1 4YF, United Kingdom

Author Note

Bo Yao https://orcid.org/0000-0003-1852-2774

No data, code, or materials were created for this theoretical paper. Preregistration was not applicable. The work was presented at the symposium 'The Sound of Thought – Exploring the Frontiers of Inner Speech' at the 24th conference of the European Society for Cognitive Psychology. It was supported by the Bial Foundation (068/2022) and an APEX Award (APX\R1\241142) funded by the Leverhulme Trust, the British Academy, Royal Academy of Engineering and Royal Society. I am grateful to Martin H. Fischer, Kate Cain, and three anonymous reviewers for their insightful comments on earlier drafts of this paper.

Correspondence concerning this article should be addressed to Bo Yao, Department of Psychology, Fylde College, Lancaster University, Lancaster LA1 4YF, UK, Email: b.yao1@lancaster.ac.uk.

Abstract

This paper introduces the Linguistic Active Inference Theory (LAIT), which proposes that inner speech augments the brain's predictive processes by transforming prior expectations and sensorimotor predictions to help reduce prediction error and uncertainty. By leveraging language's unique properties - its efficiency in representing sensorimotor information, its ability to extend across time and space, and its generativity in constructing novel predictions - inner speech enables predictive processes to transcend immediate experience, encoding complex sensory experiences into linguistic forms for perceptual inference, while decoding abstract goals into situated actions for active control. LAIT provides a unifying framework explaining how inner speech's diverse functions, varied phenomenology, and neurocognitive-developmental mechanisms all emerge from its augmentation of perceptual inference and active control. It posits that inner speech dynamically adapts its form and function in response to computational demands and ongoing prediction errors, to reduce the imprecision in the brain's generative model. This synthesis advances foundational theories and provides a roadmap for future research: generating novel testable hypotheses, motivating a shift towards dynamic and integrative methodologies, and opening new perspectives on related mental phenomena and the broader role of symbolic systems in cognition.

Keywords: Inner speech, linguistic active inference, grounded cognition, perceptual inference, active control

Introduction

Inner speech - the silent production of language in our minds - stands as one of the most familiar yet enigmatic experience of human consciousness (Alderson-Day & Fernyhough, 2015a; Fernyhough & Borghi, 2023). While many people report to experience this 'voice in the mind', science has struggled to articulate its fundamental nature and purpose. What exactly is inner speech, and *why* do we speak to ourselves in our minds?

Recent decades have seen remarkable advances in understanding inner speech's manifestations and significance. Phenomenologically, it exhibits rich diversity - from expanded dialogues to condensed phrases, experienced in one's own voice or occasionally others' (McCarthy-Jones & Fernyhough, 2011). Cognitively, inner speech serves numerous functions: maintaining information in working memory (Baddeley, 1992; Baddeley & Hitch, 1974), supporting reading and comprehension (Yao & Scheepers, 2011, 2018), facilitating problem-solving and planning (Baldo et al., 2005; D'Argembeau et al., 2011), enabling self-regulation (Fernyhough, 1996; Vygotsky, 1987), and enhancing self-awareness through reflection (Morin, 2005, 2018). Disruptions in inner speech have been linked to mental health conditions like schizophrenia and anxiety, underscoring its crucial role in psychological wellbeing (Alderson-Day et al., 2018).

These discoveries have proliferated in a multitude of theoretical frameworks: developmental accounts of how children internalise social dialogue into private self-talk (Vygotsky, 1934/1987), working memory architecture incorporating an inner rehearsal component (Baddeley, 1992), and neurocognitive models addressing multiple aspects of inner speech - from production mechanisms governing speech generation and monitoring (Carruthers, 2018; Grandchamp et al., 2019), to corollary discharge processes creating internal speech sounds (Jack et al., 2019; Scott, 2013), to perceptual mechanisms involving speech memory simulation (Tian et al., 2016; Yao et al., 2011), to prosodic features shaping thought structure (Kreiner & Eviatar, 2024).

While this theoretical proliferation reflects the field's growing vitality, it has paradoxically made it harder to grasp inner speech's essential nature. Each account captures a distinct aspect - developmental trajectory, cognitive architecture, experiential quality, or neural implementation. Yet,

integrating these perspectives is necessary to synthesise a deeper understanding of inner speech's fundamental nature.

Addressing this challenge requires stepping back from specialised investigations to examine inner speech at what Marr (1982) termed the computational level of analysis. In his framework, Marr proposed that any information-processing system can be understood at three distinct levels. The highest, computational level defines the overarching goal - what problem the system is solving and why. The middle, algorithmic level specifies the procedures to achieve that goal. The lowest, implementational level describes the 'hardware' (e.g., neural circuits) that implement these procedures. While existing research has made substantial progress at the algorithmic level (cognitive processes) and the implementational level (neural mechanisms), we have yet to fully address the fundamental question: *What is inner speech's computational purpose?*

This paper introduces the Linguistic Active Inference Theory (LAIT) to answer this question, proposing that inner speech augments predictive processes by transforming prior expectations and the resulting sensorimotor predictions, thereby contributing to the reduction of prediction errors (mismatches between predicted and observed states) and uncertainty (imprecision in the brain's generative model that gives rise to such errors), for perceptual inference and active control.

LAIT's scope centres on the self-directed deployment of internal language to modulate priors and predictions, while unifying its diverse functional, phenomenological, and neurocognitive-developmental manifestations. It does not attempt to explain communicative language use, which may follow active inference principles but requires distinct mechanistic specifications focused on shared understanding and interpersonal coordination rather than self-directed cognitive augmentation.

The paper unfolds this theoretical synthesis in three parts. First, I establish the foundations by synthesising active inference principles with grounded language processing, illustrating how these frameworks naturally converge in linguistic active inference, where inner speech augments predictive processes to support inference and control. Second, I reveal how language's unique properties - its efficiency in encoding complex sensorimotor experiences, its extendibility across

time and space, and its generativity in constructing novel predictions - transform active inference through inner speech, and how LAIT integrates the diverse observations about inner speech's functions, phenomenology, and models under a unified framework. Finally, I translate LAIT's theoretical advances into a research roadmap, formulating testable hypotheses about inner speech dynamics, outlining necessary methodological innovations to test them, and exploring applications to related mental phenomena and the broader role of symbolic systems in cognition.

PART 1: The Synthesis of LAIT

LAIT conceptualises inner speech as a linguistic augmentation of active inference, where internal linguistic processes transform prior expectations and the resulting sensorimotor predictions, thereby contributing to prediction error and uncertainty reduction for perceptual inference and active control.

I begin by introducing active inference as a fundamental principle of adaptive behaviour. Next, I explore the grounding of language in perception and action, setting the stage for incorporating inner speech within the active inference framework. Finally, I propose linguistic active inference as a synthesis of these frameworks, offering a novel perspective on inner speech as linguistic processes that actively shape perception, guide action, and transform active inference.

Cognition and Behaviour through Active Inference

Founded on the Free Energy Principle, active inference is a conceptual framework describing how living systems adaptively interact with the world by resisting a natural tendency towards disorder (Friston, 2009, 2010).

The Free Energy Principle proposes that all self-organising biological systems work to minimise an information-theoretic quantity called 'free energy' – representing the divergence between an organism's internal model of the world – called 'the generative model' as they actively generate predictions about the causes of sensory input – and the observed state of the world. This minimisation constitutes both refining the generative model to better explain observations, and acting upon the world to generate expected sensory outcomes.

Active inference operationalises the Free Energy Principle through a dynamic cycle of perception and action, where the brain continuously generates predictions and acts to minimise discrepancies between expected and actual sensory inputs (Friston et al., 2016; Pezzulo et al., 2024). The brain encodes its understanding of the world in a hierarchical generative model – a sophisticated representation of world states, dynamics, and causal structures that govern sensory observations and action possibilities.

This generative model produces hierarchical prior expectations - probabilistic hypotheses about hidden world states (prior beliefs), desired outcomes (prior preferences), and intended actions (prior policies). These priors continuously generate sensory predictions that are compared against incoming sensory inputs. When discrepancies occur between expected and actual inputs, the brain minimises these prediction errors either by updating its beliefs about the causes of observations (perceptual inference), or by acting to produce expected sensory outcomes (active control).

The degree to which prediction errors drive learning depends on their precision weights – the brain's estimate of sensory precision relative to prior precision[1]. Strongly weighted errors from precise sensory evidence drive substantial belief updating, progressively reducing uncertainty in the generative model. Weakly weighted errors from imprecise inputs are downweighted, preventing overreaction to noise.

Notably, the brain's drive to minimise prediction errors is not a mindless avoidance of novelty. Active inference reflects a sophisticated balancing act between certainty and exploration. The brain continuously seeks to reduce immediate prediction errors, yet it also strategically seeks out informative (epistemic) prediction errors that likely support learning, aiming to reduce long-term uncertainty in its generative model (Friston et al., 2017; Parr & Friston, 2019). This balance optimises evolutionary fitness by ensuring organisms neither retreat to predictable 'dark rooms' to

[1] The precision weight, or learning rate (i.e., the extent to which a prediction error updates a belief) can be expressed formally as: Sensory Precision / (Prior Precision + Sensory Precision). This ensures that the updated belief is a precision-weighted average of the prior and the sensory evidence. If sensory evidence is highly precise and the prior is imprecise, the belief shifts strongly towards the evidence, and vice versa.

avoid any prediction errors nor engage in chaotic exploration without useful model-building (Schwartenbeck et al., 2019).

This balancing act is made possible by active inference operating across multiple hierarchical levels in the brain, integrating lower-level sensorimotor predictions with higher-level conceptual and motivational beliefs to enable prediction and control across temporal and spatial scales - from immediate environmental interactions to long-term planning and strategy formation (Friston, 2008; Kiebel et al., 2008).

Active inference has been successfully applied to various cognitive and neurobiological phenomena, including perception (Parr et al., 2019), motor control (Adams et al., 2013), attention (Mirza et al., 2019), decision-making (Constant et al., 2019), and psychiatric disorders thought to stem from maladaptive predictive processes (Benrimoh et al., 2018).

Beyond these applications in basic cognition and clinical domains, examining how active inference operates at the linguistic level offers exciting insights into how language processes, like inner speech, transform predictive processes in the brain. To fully understand this linguistic active inference framework, we must next examine how language is inherently grounded within our perceptual and motor systems, providing the physical substrate through which active inference manifests in cognition.

The Grounding of Language in Perception and Action

Grounded cognition posits that cognitive processes, including language, are fundamentally rooted in the body's interactions with the world (Wilson, 2002), providing a crucial bridge between language processes and sensorimotor active inference.

This theoretical framework emphasises how linguistic representations derive from grounded experiences, establishing experiential foundations for meaning that develop through continuous learning and refinement throughout life (Barsalou, 1999, 2008). While not every experience has a corresponding linguistic representation, many experiences are routinely coupled with words, phrases, or other linguistic expressions. This established coupling results in automatic co-activation: perceiving an actual cat partially reactivates the linguistic label 'cat', and conversely,

hearing or thinking the word ‘cat’ partially reactivates sensorimotor experiences associated with cats. This automatic bidirectional relationship allows language to effectively ‘encode’ complex sensorimotor patterns into compact linguistic forms, while these linguistic forms can be deliberately ‘decoded’ through mental simulations of situated sensorimotor experiences. This linguistic encoding-decoding mechanism provides the foundation for language’s role in predicting and interpreting sensory experiences and translating abstract goals into situated bodily actions.

Empirical evidence strongly supports this grounded view of language. Multiple studies demonstrate that language comprehension actively interacts with and influences concurrent perceptual and motor tasks (Glenberg & Kaschak, 2002; Stanfield & Zwaan, 2001; Zwaan & Taylor, 2006), while perceptual and motor states reciprocally influence language production (Casasanto & Chrysikou, 2011; Hostetter & Alibali, 2008; Pouw et al., 2020). This bidirectional relationship extends to the neural level, as neuroimaging research reveals that processing sensory and action-related words recruits neural substrates typically involved in actual perception and action (Kiefer et al., 2008; Pulvermüller, 2005, 2013), while pre-activation of sensorimotor brain regions using transcranial magnetic stimulation facilitates subsequent language processing (Pulvermüller et al., 2005; Willems et al., 2011).

These findings demonstrate how language processes are intrinsically integrated with the brain’s sensorimotor systems, suggesting that inner speech might operate through similar grounded mechanisms within an active inference framework.

Inner Speech as Linguistic Active Inference

By synthesising active inference with grounded language processing, linguistic active inference emerges as a framework for understanding how the brain deploys language internally to support its imperative to reduction uncertainty (imprecision) in its generative model via precision-weighted prediction error minimisation. At its core, LAIT posits that inner speech dynamically transforms both the content and precision of priors and the sensorimotor predictions they generate, thereby augmenting active inference’s fundamental mechanisms – constructing and refining predictions, interpreting prediction errors, formulating causal relationships, and planning actions. In

hearing individuals, these internal linguistic processes often encompass articulatory, phonological, and prosodic representations of spoken language, which we consciously recognise as 'inner speech'.

Before we explore how linguistic active inference may unfold, it is important to clarify the relationship between the brain's automatic predictive processes and the conscious experience of inner speech. While many linguistic predictions can influence cognition rapidly and unconsciously, they manifest as conscious inner speech when prediction errors are too significant, persistent, or complex to be resolved automatically by specialised local systems. When prediction errors exceed localised processing capacity, they trigger a transition from automatic, unconscious processes to deliberate, conscious problem-solving - a shift requiring global integration across diverse cognitive operations. Inner speech, as the conscious manifestation of linguistic predictions, serves as an ideal mechanism for this cognitive transition. It 'broadcasts' unresolved prediction errors while unifying crossmodal information – from sensory observations and abstract concepts to situational knowledge and motor commands – into a shared, temporally and causally structured format that enables coherent, increasingly complex inference. This linguistic broadcasting function aligns conceptually with Global Workspace theories of consciousness (Baars, 1997; Dehaene et al., 2014), which propose that consciousness emerges when local processing limitations necessitate global information integration.

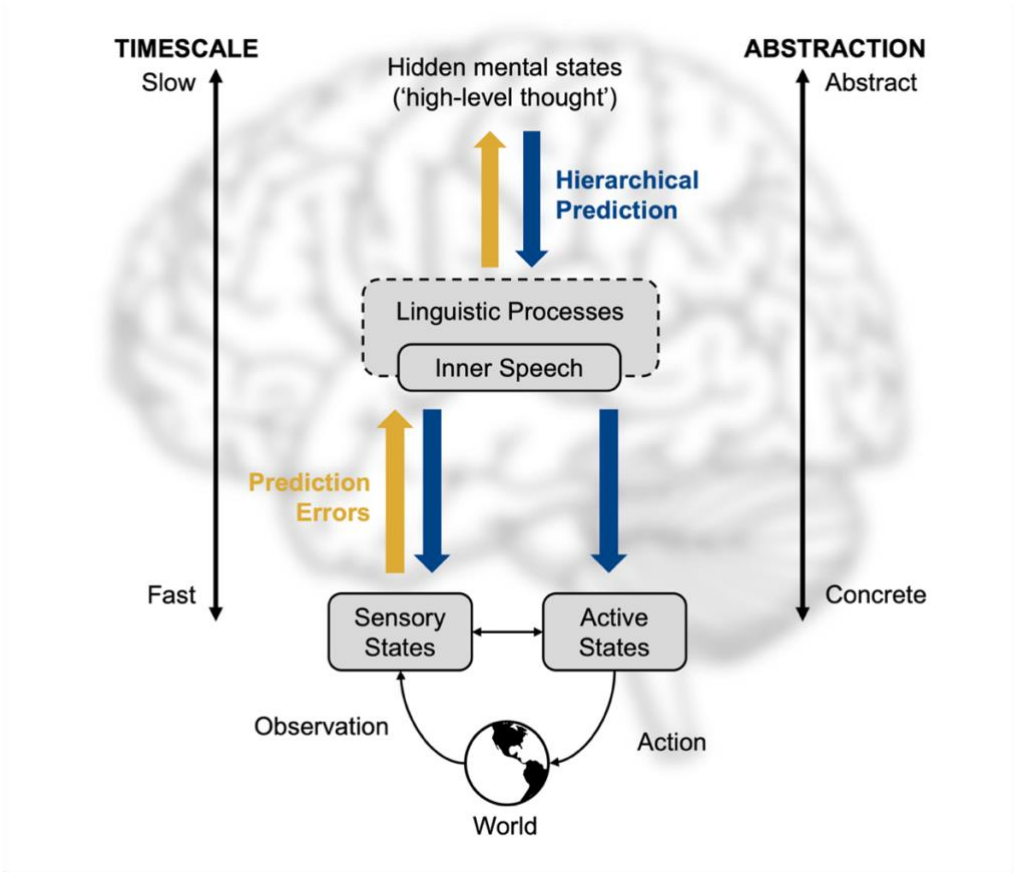
To understand how inner speech augments predictive processes in perception and action, we must examine the hierarchical architecture in which it is embedded. As illustrated in **Figure 1**, this architecture addresses the challenge of connecting the brain's higher-order, slow-changing goals (e.g., 'survival') with the high-dimensional, fast-changing data of the sensorimotor periphery (e.g., raw visual input, precise muscle commands). That is, the top levels of the hierarchy are too abstract to directly encode sensations and guide action, while the bottom levels are too detailed and complex to represent goals and intentions. This creates a representational mismatch, demanding an intermediary mechanism that can bridge these incompatible representations.

Inner speech operates as the crucial computational intermediary in this hierarchy, functioning as a 'sweet spot' that efficiently translates between abstract goals and concrete

sensorimotor states. In the top-down (decoding) direction, it transforms high-level goals into structured linguistic plans (*"Check over there"*), which then generate situated sensorimotor predictions to guide actions (e.g., getting up from the chair, and walking to the window on the far side of the office to look down at the street). Conversely, in the bottom-up (encoding) direction, it distils noisy, complex prediction errors into compact linguistic representations of their causes (*"That's not Geroge!"* or *"Raining!"*) that can efficiently update higher-level models without overwhelming them with raw detail. This bidirectional transformation process provides a bridge for abstract goals to cascade down to modulate sensorimotor predictions whilst prediction errors percolate up to inform goal pursuit.

Figure 1

Hierarchical architecture of linguistic active inference



Note: This architecture depicts hidden mental states (representing abstract, high-level thought such as goals, intentions, and high-level beliefs) at the top level, which are decoded into linguistic forms through the linguistic processes in the middle layer. These processes, which can operate unconsciously, are consciously experienced as inner speech - the internal production and experience of language - which modulates top-down prior expectations and the resulting sensorimotor predictions in lower-level sensory states (neural representations of perceptual information derived from environmental observations) and active states (neural representations driving physical actions and interactions with the world) that interface directly with our environment. The entire system functions along gradients of abstraction (from concrete sensorimotor representations to abstract conceptual knowledge) and processing timescales (from fast sensorimotor operations to slower conceptual thought processes). Information flows bidirectionally through this hierarchy: top-down hierarchical predictions translate abstract mental states into linguistic forms and subsequently generate predictions about sensory inputs and actions, while bottom-up prediction errors are encoded into and resolved through linguistic reformulations that update abstract thought, driving learning and model refinement at all levels.

To illustrate how these mechanisms unfold, imagine you are working late, on a paper that is long overdue. Suddenly, you observe a long, curly shadow in the dark corner of the office, and the brain initiates perceptual inference. The brain holds prior beliefs about what one would typically encounter in an office, where such a shadow may be unexpected. While much of the low-level visual processing may happen unconsciously, this unexpected observation generates a substantial prediction error to engage conscious attention, manifesting as a spontaneous inner exclamation *"What?"* or question *"Is that a …?"*, thereby triggering more deliberate cycles of linguistic active inference.

Once attention and linguistic processes are engaged, inner speech can unfold in various ways to help the brain reduce the prediction error. It can be formulated to construct and compare competing hypotheses (*"Is that a … snake or a cable?"*), exploring multiple beliefs that could explain the observation. When committing to testing one of these beliefs, inner speech can sharpen its precision by consciously labelling it (*"Snake!"*), which in turn generates more specific sensory predictions about the diagnostic features of a snake (e.g., scaly texture, a snake-like head shape), and directs attention and eye gaze to seek confirming evidence.

If these predictions are met, the snake belief is confirmed and higher-order survival imperatives are expressed in linguistic forms for causal inference (*"Danger!"*) and action planning (*"Call security!"*), which are then decoded into situated sensorimotor contexts for active control (e.g., holding your breath, getting up slowly before backing away quietly to the office door, and run…).

Conversely, if these predictions are violated (e.g., the shadow appears to be connected to a wall socket), inner speech can explicitly dismiss this implausible belief (*"Can't possibly be a snake!"*), and amplifying a more probable alternative (*"Must be a cable"*), pivoting to a different path of sensory sampling and inference.

If none of the tested beliefs adequately explain the sensory input - perhaps because the input remains too noisy despite the reallocation of attention - inner speech may explicitly express this persistent sensory uncertainty (e.g., *"Not sure…"*) and formulate epistemic actions (*"Gonna turn the light on and take a look"*) to gather more decisive sensory data.

This scenario illustrates how inner speech exploits the brain's established language-world couplings to modulate the content and precision of prior beliefs and the resulting sensorimotor predictions for active inference.

Indeed, a growing body of empirical research supports this language-world interaction, demonstrating that concurrent linguistic processing modulates perceptual and motor functioning. When participants visually searched for the numeral 2 amongst 5s, hearing *"find the two"* or *"ignore fives"* immediately before searching improved response times and search efficiency (Lupyan, 2007). This mirrors our earlier example where inner speech explicitly labelled a prior belief (*"Snake!"*) or suppressed it (*"Can't possibly be snake!"*), thereby modulating the precision of prior beliefs and the resulting sensory predictions for directing attention and eye gaze.

The linguistic enhancement of sensory predictions extends beyond simple attentional management to fundamentally alter perceptual sensitivity. Hearing target words like *"chair"* significantly enhances the detection of otherwise hard-to-detect visual objects (Lupyan & Ward, 2013). Crucially, these linguistic labels meaningfully change neural activity in the visual cortex, occurring within 100ms with effects that vary predictably based on the match between stimulus shape and the shape denoted by the label (Noorman et al., 2018). This suggests that linguistic labelling strengthens prior beliefs that translate into more precise lower-level sensory predictions incorporating specific visual features, making detection more likely from uncertain, imprecise sensory input where perceptual inference is dominated by more precise prior beliefs.

Like inner speech, these linguistic labels can be self-generated to influence perception. Overt naming of targets during visual search facilitates performance when the 'imagery concordance' between the names and visual targets is high and impairs performance when concordance is low (Lupyan & Swingley, 2012). This demonstrates that self-generated linguistic formulations - similar to the inner speech in our office shadow example - deploy predictive models that generate visual predictions. Matching visual predictions facilitate visual recognition whereas mismatching predictions cause predictions errors that slow down performance.

This linguistic modulation extends to action as well. Action words activate motor cortex during word production (Oliveri et al., 2004), and hearing action words like *"squeeze"* while

observing corresponding actions increases both visual attention to target objects and motor cortex excitability (Franklin et al., 2020). Moreover, when participants observe novel, sequential actions for later reproduction, disrupting inner speech during observation significantly impairs the encoding and subsequent recall and reproduction (Gervasi et al., 2025). These findings indicate that linguistic formulations of actions - whether heard or self-generated - can modulate action-related prior expectations that encompass both motor predictions and associated sensory consequences.

Further evidence for inner speech-like linguistic active inference comes from studies where participants perform perceptual tasks silently. Russian speakers, who have distinct categorical terms for light blue (‘goluboy’) and dark blue (‘siny’), discriminate these colours faster than English speakers (Winawer et al., 2007). Crucially, this advantage disappears under verbal interference but not spatial interference, suggesting that participants likely engage linguistically-mediated categorical perception during colour discrimination, and that disrupting this inner linguistic process eliminates the enhanced perceptual sensitivity.

Most importantly, Cibelli et al. (2016) provided a formal computational account of these language-specific effects on colour representations, which aligns closely with the proposed LAIT mechanics. They modelled colour memory as the convolution of two probability distributions: continuous, non-linguistic colour representations and categorical, linguistic colour representations. Categories most strongly affect colour memory when perceptual information is uncertain - precisely the conditions where linguistic formulations would be most impactful when modulating prior probabilistic distributions during active inference. The model demonstrates how categorical linguistic representations modify non-linguistic representations through probabilistic convolution, with precision weights determining whether linguistic or perceptual information dominates the resulting colour representation. This precision-weighting mechanism mirrors the process where inner speech modulates the influence of top-down prior beliefs and sensory predictions relative to bottom-up sensory evidence.

While this empirical foundation demonstrates that linguistic processes can modulate sensorimotor processing in theoretically predicted ways, a significant gap remains between laboratory findings and the complexity of natural inner speech. These studies predominantly

examine the effects of single words or simple phrases on simple perceptual and motor tasks. Inner speech, by contrast, exhibits rich phenomenological diversity - from fragmented verbal thoughts to expanded internal dialogues - and influences complex cognitive functions including reasoning, planning, and self-regulation. The existing literature, while supportive, captures only a fraction of inner speech's potential computational contributions. This raises a fundamental question: What computational advantages does language provide that non-linguistic active inference cannot achieve? Understanding these advantages is crucial for grasping why inner speech has evolved as a cognitive tool for thinking and how it transforms human cognition.

PART 2A: Why Use Language for Active Inference?

While other representational systems could theoretically serve as intermediaries between high-level abstract states and low-level sensorimotor states, inner speech provides three key properties of language that uniquely transform active inference: efficiency, extendibility, and generativity.

Efficiency

As previously discussed, language's contribution to efficient active inference lies in its capacity as a grounded symbolic system that encodes high-dimensional sensorimotor experiences into compact, manipulable linguistic forms (Barsalou, 1999, 2008). These linguistic forms function as cognitive shortcuts, indexing complex multisensory predictions without simulating each sensory detail in full (Connell, 2019). This efficiency appears fundamentally designed into language through evolutionary pressures (Gibson et al., 2019), with language acting as a high-level control system for manipulating mental representations (Lupyan, 2016). This property makes inner speech particularly relevant to cognitive domains requiring rapid categorisation and schema-based prediction, such as object recognition or decision-making in routine scenarios, where computational overhead can be minimised without sacrificing predictive accuracy.

Evidence for these shortcuts spans development and adulthood. From infancy, language shapes object categorisation (Ferguson & Waxman, 2017), with infants forming label-mediated

conceptual understandings by their first year (Westermann & Mareschal, 2014), and caregivers facilitating categorisation through naming and linguistic marking (Gelman & Meyer, 2011; Waxman, 2013). In adulthood, verbal labels provide uniquely efficient access to conceptual information compared to nonverbal cues (Lupyan & Thompson-Schill, 2012), rapidly modulating cognition and perception (Lupyan, 2012), and facilitating both category formation and learning (Lupyan, 2006; Lupyan et al., 2007; Perry & Lupyan, 2014). Cross-cultural evidence further demonstrates how linguistic differences in number systems and colour terms shape categorical perception (Gordon, 2004; Roberson et al., 2005), with language offering distinct indicators of category membership even when relevance isn't explicit (Gervits et al., 2023).

Beyond simple categories, language facilitates the formation and access of schemas - complex knowledge structures that guide processing across recurring scenarios (Bartlett, 1932). These schemas (e.g., the 'restaurant script') efficiently package knowledge about typical sequences, roles, and expectations (Schank & Abelson, 1977), allowing rapid generation of situation-appropriate predictions and often filling missing details without computing each element anew. Crucially, these schemas are grounded in the physical environment through a causal-predictive cycle (Roy, 2005), functioning as situated concepts that support contextual simulations (Barsalou, 2009). Through repeated experience, schemas become increasingly refined predictive models that facilitate rapid interpretation and response to familiar situations (Mandler, 1984; McClelland & Rumelhart, 1999), functioning essentially as pre-compiled predictive models accessible through inner speech.

Collectively, this evidence demonstrates how linguistic representations enhance efficiency by transforming complex, high-dimensional sensorimotor experiences into compact, manipulable shortcuts that reduce computational overheads during inference, while retaining access to the rich sensorimotor knowledge when needed.

Crucially, language's efficiency extends beyond these cognitive shortcuts to encompass a fundamental transformation in how learning operates. Unlike statistical learning that gradually adjusts continuous probability distributions based on accumulated perceptual patterns, linguistic representations deploy prior knowledge that imposes probabilistic constraints that 'warp'

continuous sensory information into more categorical representations (Goldstone & Hendrickson, 2010; Harnad, 1987; Lupyan, 2012). This linguistic 'warping' function dramatically increases the precision of deployed predictive models whilst suppressing alternatives, creating 'high-stakes' inference cycles where the precise priors either receive confirming evidence, leading to confident model confirmation, or generates larger prediction errors that drive faster belief updating (Clark, 2013; Friston et al., 2017). This categorical transformation effectively accelerates learning by transforming gradual statistical adjustments into more decisive updates, thereby expediting the refinement of the brain's internal models, as compared to purely data-driven statistical learning (Kemp et al., 2007).

Extendibility

Language significantly extends the scope of active inference beyond immediate sensorimotor experiences, enabling inner speech to construct predictive models that span across time, space, and levels of abstraction. This makes inner speech particularly relevant in domains involving displaced or abstract reasoning, which cannot be adequately addressed by sensorimotor active inference alone.

Displacement provides the foundational property enabling language's extendibility, allowing the representation of things not immediately present in the sensory environment (Hockett, 1960; Zwaan, 2014). Displacement directly enhances active inference by enabling predictions about entities beyond immediate perception, as evidenced by studies showing how language influences perceptual uncertainty. Linguistic labels can enhance visual awareness of otherwise invisible objects (Lupyan & Ward, 2013), facilitate categorical perception of unfamiliar faces (Kikutani et al., 2008), and affect early visual processing for blurred images and unfamiliar objects (Abdel Rahman & Sommer, 2008; Weller et al., 2019). These findings illustrate how inner speech shapes predictive models and guides attention when sensory input is ambiguous or unavailable - a crucial capability for navigating uncertain environments.

Language's extendibility reaches its fullest expression when enabling active inference about abstract concepts lacking direct sensory mappings. For abstract conceptual domains, language

provides structure and grounding that sensorimotor simulation alone cannot achieve. The Words As Social Tools perspective (Borghi et al., 2019) and related frameworks (Dove, 2018; Dove et al., 2022) emphasise language's crucial role in abstract concept acquisition and processing. Empirical evidence demonstrates language's involvement in abstract concept acquisition and representation (Fini et al., 2022; Granito et al., 2015). Similarly, language extends active inference into social domains by enabling predictions about others' mental states - with early language skills predicting later Theory of Mind performance (Astington & Jenkins, 1999), training on sentential complements improving Theory of Mind (ToM) (Hale & Tager-Flusberg, 2003), and evidence suggesting a bidirectional relationship between language and social understanding (Slade & Ruffman, 2005; Stanzione & Schick, 2014).

Building upon these various forms of extendibility, inner speech transforms immediate sensorimotor active inference into linguistically-extended prediction across time, space, and conceptual domains, allowing us to navigate increasingly complex forms of uncertainty.

Generativity

While efficiency benefits from linguistic encoding and extendibility elevates predictions beyond immediate experience, language's inherent generativity enables inner speech to construct novel and increasingly complex predictive models for active inference. This generative capacity makes inner speech particularly valuable for abstract reasoning and creative problem-solving, enabling us to construct innovative mental models when navigating scenarios that transcend our learned sensorimotor knowledge.

Language's combinatorial nature enables inner speech to generate novel predictions through systematic mappings across domains. Metaphoric thinking exemplifies this process - studies show how spatial concepts structure temporal reasoning (Casasanto & Boroditsky, 2008; Gentner et al., 2002), with learning new metaphors reshaping mental representations of time (Boroditsky, 2018). Similar mappings through physical metaphors extend to mathematical understanding (Lakoff & Núñez, 2000) and concepts related to emotion (Crawford, 2009) and abstract size (Yao et al., 2022). Beyond metaphor, combining linguistic concepts generates

emergent meanings beyond component parts (Estes & Ward, 2002), as seen in scientific discourse where novel terms create new predictive frameworks (Pecman, 2014). Through this combinatorial capacity, inner speech enables predictions impossible through direct sensorimotor simulation alone.

The recursive structure of language further enhances generativity by enabling hierarchical organisation of predictions. This recursion allows inner speech to generate nested models operating at multiple abstraction levels simultaneously (Chomsky, 1965; Johnson-Laird et al., 2022) - critical for complex problem-solving and abstract reasoning. Evidence for this ranges from impaired nonverbal problem-solving in individuals with limited early language exposure (Baldo et al., 2015), the centrality of nested structures in scientific hypothesis formation (Nersessian, 2008), to superior performance of AI agents using hierarchical language (Hu et al., 2019). Additionally, while the study of acquired aphasia is complex and does not necessarily imply a loss of inner speech (Fernyhough & Borghi, 2023), associated impairments in nonverbal problem-solving (Baldo et al., 2015) and hypothesis generation (Varley, 2002) nonetheless highlight that the language system is integral for building complex inferential structures beyond immediate experience.

The socio-cultural dimension of language extends generativity beyond individual cognition through transmission across communities and time. Through language, humans access others' predictive models - readers construct detailed situation models from text (Johnson-Laird, 1983; Zwaan & Radvansky, 1998), enabling social-cognitive predictions without direct experience (Mar & Oatley, 2008). In cultural learning, language accumulates predictive knowledge across generations (Csibra & Gergely, 2009; Tomasello, 2019), with distinct linguistic traditions shaping different predictive frameworks (Gentner & Goldin-Meadow, 2003; Kirby et al., 2008) and cultural transmission refining collective models (Boyd et al., 2011). Inner speech thus provides access to this shared repository of predictive knowledge, dramatically extending active inference beyond personal experience.

Taken together, these three linguistic properties - efficiency, extendibility, and generativity - enable inner speech to exploit language's computational advantages, facilitating rapid construction

and manipulation of predictive models across time and abstraction, as well as access to culturally transmitted predictive frameworks, thereby augmenting active inference's scope and capabilities.

PART 2B: How Does Linguistic Active Inference Unify Diverse Functions, Phenomenology, and Models of Inner Speech?

Understanding how inner speech augments active inference's core computational mechanisms – perceptual inference and active control – reveals how its diverse cognitive functions, phenomenological qualities, and neurocognitive-developmental mechanisms emerge from these mechanisms to support active inference's overarching aim of reducing prediction errors and uncertainty.

Unifying Diverse Functions of Inner Speech

Inner speech serves diverse cognitive functions, from directing attention and supporting working memory to enabling categorisation, planning and problem-solving, as well as self-reflection and regulation. These functions emerge from how inner speech augments linguistic active inference - using language to interpret sensory information and control actions for prediction error and uncertainty reduction. These processes work together in continuous perception-action cycles that manifest across various cognitive domains.

Inner speech for perceptual inference

Inner speech supports perceptual inference by formulating precise prior beliefs about the causes underlying our observations, which then generate specific, diagnostic sensory predictions that are tested against incoming sensory information. When sensory input is ambiguous or 'imprecise', prediction errors are downweighted due to the comparatively lower reliability of the sensory evidence relative to more precise priors, which remain unchallenged, effectively 'warping' ambiguous sensory data within linguistically formulated conceptual frameworks for perception. Conversely, when sensory input is precise and reliable, prediction errors carry greater weight,

driving the updating and refinement of prior beliefs through error minimisation, progressively reducing 'epistemic' uncertainty about world states over time.

For example, seeing an ambiguous furry silhouette might trigger inner speech (*"Cat?"*), which then generates feline-specific sensory predictions that are tested against the sensory data. When the input is highly imprecise and no clear evidence contradicts this belief, the perception of a cat is sustained. However, should specific, diagnostic sensory evidence emerge, such as an unusually large, fluffy tail that contradicts these feline predictions, inner speech may help update the existing belief (*"That's called a Maine Coon? I didn't know some cats can have bushy tails…"*), thereby refining the conceptual boundaries of what 'cat' entails. Alternatively, inner speech may help reduce this prediction error by entertaining an alternative causal hypothesis (*"…or a fox?"*), generating alternative predictions to better match the sensory input. This linguistic formulation of 'perceptual hypotheses' operates across diverse perceptual domains. In social contexts, for instance, observing someone's frowning expression might similarly trigger inner speech (*"Are they upset? Is it [to do with] me?"*), generating prior predictions about the underlying causes and subsequent behavioural expectations.

The proposed prior modulation mechanism is supported by evidence indicating that inner speech shapes aspects of nonverbal perception (Lupyan et al., 2020). As previously noted, Russian speakers, with distinct terms 'goluboy' for light blue and 'siny' for dark blue, exhibit better discrimination between these hues than English speakers, who use the single term 'blue' (Winawer et al., 2007). Similarly, Mongolian speakers, with separate terms for light blue ('qinker') and dark blue ('huhe'), show faster sorting and visual search than Chinese speakers using one term (He et al., 2019). The advantages in colour perception among Russian and Mongolian speakers are diminished by verbal interference, highlighting covert linguistic labels are likely used to modulate perceptual inference of subtle colour distinctions. Moreover, these cross-language differences in colour perception are particularly pronounced under perceptual uncertainty, as demonstrated by probabilistic models integrating universal perceptual spaces with language-specific categories (Cibelli et al., 2016). Beyond colour perception, simply naming an object enhances its visual

characteristics, aiding in categorisation and recognition across age groups, from infants to adults (Landau & Leyton, 1999; LaTourrette et al., 2023).

Language's involvement in perception is further supported by neuroimaging and electrophysiological studies. Ambiguous visual stimuli activate language-related brain regions, such as the left inferior frontal gyrus (L-IFG) (Bar et al., 2001). Hearing words enhances early visual processing more than nonverbal sounds when recognising familiar animals and artifacts, as indicated by increased P1 event-related potential (ERP) component (Rabovsky et al., 2012). Although the latter findings concern externally driven language processing, the underlying coupling between language and perceptual experience could be drawn upon by inner speech to exert similar perceptual influences during active inference.

Moreover, inner speech's influence over perception becomes particularly pronounced when dealing with complex causal models in domains like social cognition and emotion, where sensory evidence alone provides insufficient information to resolve uncertainty. In these contexts, linguistic processes specify hypotheses about hidden causes, generating predictions that can be tested against the limited observations available. For example, in social interactions, dialogic inner speech may maintain and test alternative hypotheses about others' mental states (*"is she scared? or just calm?"*) that generate sensory predictions about facial expressions, vocal patterns, and body language to explain the limited observations (Fernyhough, 2016). Similarly, in emotional processing, linguistic labels like 'anxiety' help coordinate predictions across multiple perceptual channels - from interpreting others' behavioural cues in social situations to recognising patterns in one's own interoceptive signals. Through continuous cycles of prediction and error correction, inner speech helps transform complex, multimodal sensory patterns into meaningful psychological and emotional understanding, forming the basis for subsequent emotion expression and regulation (Alderson-Day & Fernyhough, 2015a; Kittani & Brinthaupt, 2024).

In sum, inner speech's role in perceptual inference helps explain its diverse cognitive functions - from categorisation to problem-solving and self-reflection. By providing linguistic representations for interpreting and predicting sensory experiences, inner speech shapes how we understand both current situations and anticipated states.

Inner speech for active control

Inner speech supports active control by formulating hierarchical prior expectations for desired states, from abstract goals to action plans. This linguistic formulation drives a cascade of increasingly specific priors that guide action selection and execution, whilst the fulfilment of these desired states progressively reduces prediction error between these states and a current state, as well as 'pragmatic' uncertainty about the brain's ability to maintain homeostasis and achieve intended outcomes.

For example, perceptual inference of interoceptive signals results in the belief *"Hungry!"*, which creates a large gap between the current state of hunger and the desired state of nourishment, derived from the higher-order belief that the nourishment is necessary for survival. The brain attempts to reduce this prediction error by performing actions to change the current state. Initially, the brain formulates an imprecise action goal such as *"make dinner"*, which prompts epistemic actions to gather sensory data, such as checking the fridge or examining saved receipts. Perceptual inference of new sensory evidence (e.g., seeing chicken and curry paste) leads to the formulation of a more specific prior policy - *"Thai green curry it is!"*. This more specific policy is then decoded into a hierarchy of increasingly granular action priors (e.g., *"slice the baby corn into strips"*), each driving specific motor actions to alter the world (baby corn) into the desired state (strips) through the action-perception feedback loop. Through this hierarchical cascade, each prior for a desired state is progressively fulfilled from the most granular level upward, until hunger transforms into nourishment, thereby reducing 'pragmatic' uncertainty about achieving desired outcomes.

Inner speech influences active control across multiple levels of organisation, from basic attentional guidance to complex self-regulatory strategies. At the basic level, inner speech supports active control through verbal mediation of attention and behaviour. When faced with perceptual ambiguity, such as seeing a furry silhouette, inner speech (*"cat?"*) directs our gaze and attention towards features that could confirm or disconfirm our expectations. During intentional visual search, verbally rehearsing goals (*"keys, keys..."*) simulates relevant sensory features like metallic glints or key-like shapes, facilitating pattern recognition. Inner speech also maintains and updates

task rules (*"now sort by colour, not shape"*) while evaluating performance (*"that matches, nice!"*) to reinforce successful actions. Inner speech's role in these control functions is evident in how verbal interference disrupts both attentional and cognitive control (Baldo et al., 2005; Emerson & Miyake, 2003; Tullett & Inzlicht, 2010).

Building on these foundational control mechanisms, inner speech enables more sophisticated forms of self-regulation and planning. In emotional regulation, inner speech helps manage interoceptive prediction errors. For example, using emotion labels to categorise a volatile interoceptive state can produce a more precise, distinct predictive model (posterior) for selecting subsequent action policies for regulation (Barrett, 2017). Linguistic distancing - using third-person pronouns or one's own name in self-reflection – may decouple the emotion prior (third-person) from the interoceptive input (first-person), thereby creating a prior-dominant, more stable posterior of the emotional state, facilitating the exploration and selection of regulatory strategies (Kross et al., 2014; Orvell et al., 2021). Such strategies are more effectively formulated as precise if-then plans (e.g., *"If I see blood, then I will stay calm"*). These specific plans encode precise triggers and action policies, thereby reducing emotional reactivity – closely linked to the magnitude of interoceptive prediction error (Seth, 2013) - more effectively than vague intentions (e.g., *"I will not be disgusted"*), as confirmed by electroencephalographic (EEG) measures showing altered neural responses to emotional stimuli (Gallo et al., 2009).

This capacity for formulating structured action policies to reduce prospective prediction errors extends to complex sequential planning, where inner speech decomposes abstract goals into manageable, more precise steps. Studies using the Tower of London task demonstrate that disrupting inner speech through articulatory suppression - a verbal interference manipulation widely used to investigate the role of language in cognition (Nedergaard et al., 2023) - significantly impairs performance (Lidstone et al., 2010), highlighting how verbal self-instruction supports systematic action planning under uncertainty. Consistent with these findings, research involving participants building toy models from memory reveals impaired performance when inner speech is disrupted through articulatory suppression, demonstrating that inner speech enhances event memory by making sequential representations more efficient (Banks & Connell, 2024). Similarly,

when participants observe novel, sequential actions for later reproduction, disrupting inner speech during observation significantly impairs the encoding and subsequent recall and reproduction, further confirming inner speech's facilitatory role in encoding and planning complex action movements (Gervasi et al., 2025). Recent advances in robotics validate this planning function, showing that natural language feedback improves artificial systems' ability to plan and execute complex sequential tasks in embodied environments (Huang et al., 2022).

Inner speech also supports epistemic actions which reduce uncertainty through active exploration and information seeking. These functions of verbal mediation can be observed developmentally through *private speech* - the audible self-directed speech or self-talk that young children produce during activities before it becomes fully internalised as inner speech (Vygotsky, 1934/1987; Winsler et al., 2009). Both private speech and inner speech are self-directed rather than communicative, supporting active inference within the self rather than establishing shared understanding with others. This self-directed nature allows for semantically dense language incorporating personalised contexts, as opposed to more explicit, mutually accessible language required for shared understanding between interlocutors. Private speech provides audible linguistic mediation that takes more expanded forms and structure (e.g., fully grammatical self-directed statements like *"Should I double-check if this will work?"*, referring to personalised referents that only the self understands). Through development, it gradually transitions to inner speech which can take either expanded forms similar to private speech or highly condensed forms (e.g., *"double-check?"*) optimised for rapid active inference (Fernyhough, 2004).

Studies show that the internalisation of private speech from preschool to first-grade coincides with significant development in self-directed questioning (*"I wonder why..."*, *"What if..."*), suggesting that expanded forms of inner speech may similarly employ questioning structures to help coordinate exploratory behaviour and curiosity-driven learning through linguistic scaffolding of hypothesis generation and testing (Jirout & Klahr, 2020; Ronfard et al., 2018).

At higher levels of control, inner speech guides goal pursuit and motivation maintenance by formulating and refining policies to minimise prospective prediction error. Research reveals that the form of inner speech significantly influences goal pursuit - for instance, interrogative self-talk (*"Will*

I?") enhances motivation and performance compared to declarative statements (Senay et al., 2010); from an active inference perspective, an interrogative prompt initiates an epistemic process, simulating potential outcomes and one's capacity to succeed, thereby reducing uncertainty about the best policy to adopt. A simple declarative statement (*"I will"*) may represent a less-scrutinised, and therefore more error-prone, prior. Given the developmental continuity and structural similarities between private speech and expanded inner speech, these findings suggest that the latter could likewise employ interrogative forms for goal regulation. This goal regulation function is supported by computational modelling, demonstrating how inner speech enhances cognitive flexibility in goal-directed tasks (Granato et al., 2020), and its reduced use may underpin the cognitive flexibility challenges observed in autism (Granato et al., 2022). Through conscious self-reflection, inner speech enables adjustment of future action plans, helping to refine strategies based on previous outcomes (Morin et al., 2018).

At the interpersonal level, inner speech coordinates complex social interaction simulations by integrating multiple predictive processes. Individuals use inner speech to simulate potential social encounters (Morin et al., 2018), leveraging predictive mechanisms from spoken interactions (Pickering & Garrod, 2013; Pulvermüller & Fadiga, 2010) to generate both their own utterances and anticipated responses. This social simulation coordinates multiple levels of prediction, from immediate emotional reactions to long-term relationship dynamics, weaving together emotional self-regulation (*"stay calm"*), perspective-taking (*"they might feel defensive"*), and strategic planning (*"if I apologise first..."*) to minimise prospective social prediction errors.

In sum, inner speech's role in active control manifests through its capacity to formulate hierarchical priors, from abstract goals to precise action policies that structure behaviour across multiple levels. These control functions involve applying or exploring more precise prior policies that are more likely to minimise prospective prediction error between current and desired future states, as exemplified by how inner speech translates abstract goals into specific action sequences, attenuates interoceptive precision for distanced self-regulation, and coordinates multiple predictive processes for social navigation. This active control function critically complements inner speech's role in perceptual inference, where linguistic processes not only

interpret current states but actively shape future ones through the generation and fulfilment of hierarchical predictions.

A unified framework for inner speech functions

LAIT advances our understanding of inner speech by suggesting that its diverse functions emerge from a linguistic augmentation of active inference. Unlike previous accounts that describe each inner speech function in isolation, LAIT proposes that both basic operations (such as perceptual categorisation) and complex functions (such as emotion regulation and social reasoning) arise from shared underlying processes of perceptual inference and active control. The common computational principles explain not just what inner speech does, but *why* these specific functions evolved within a unified system - they work together to efficiently reduce prediction errors and model uncertainty across diverse cognitive domains, contexts, levels of complexity, and timescales.

Beyond this unification, LAIT additionally suggests that inner speech functions are hierarchically organised and synergised together. At the lower level, inner speech coordinates perception and action in continuous cycles: perceptual inference shapes action planning while action planning generates predictions that guide perceptual attention. These foundational processes support higher-level functions like goal pursuit and social reasoning, which in turn modulate lower-level processes through top-down predictions. This hierarchical organisation explains why disrupting inner speech often has cascading effects across multiple functions - for instance, articulatory suppression impairs both basic perceptual discrimination and complex problem-solving because it disrupts the shared linguistic predictive mechanisms that coordinate across multiple levels of active inference.

LAIT further predicts that inner speech functions should dynamically respond to prediction error signals across hierarchical levels and timescales. This process involves not only reacting to current prediction errors but, crucially, simulating action plans and future outcomes to minimise prospective prediction errors, a mechanism central to many proposed functions of inner speech. Consider emotional regulation: the process often begins with a current prediction error about an

unexpected state (*"I'm feeling anxious"*), which triggers causal modelling (*"Is it to do with the upcoming deadline?"*). This inference then informs the selection of an action policy (*"I should create a revision plan"*) based on its predicted success in minimising future error, leading to a more desirable and expected state (*"I will probably feel calmer"*). This prospective simulation capacity supports many other proposed functions. Practising for social encounters, for instance, explicitly simulates conversational policies to select responses that most effectively minimise potential social prediction errors. Creative production can be framed similarly as cycles of generating novel action policies (e.g., a line of poetry, a hypothesis) and internally simulating their perceptual consequences against a desired goal to select the next iterative steps. Even mental rehearsal involved in acquiring a new language may represent the use of internal simulations to strengthen the predictive mappings required for building generative models in the target language. This continuous shifting between resolving present uncertainty and modelling future possibilities across cycles of perceptual inference and active control constitutes how inner speech augments the brain's ability to adaptively allocate its finite cognitive resources, navigating both a complex present and an open-ended future. By revealing how these diverse functions emerge from a single set of computational mechanisms, this unified account explains both the variety and coherence of inner speech functions, providing a versatile and powerful framework for understanding how humans navigate increasingly complex cognitive challenges through linguistic active inference.

Unifying Diverse Phenomenology of Inner Speech

Accompanying its functions, inner speech exhibits rich phenomenological diversity, varying along several key dimensions including condensation, dialogicality, voice qualities, and spontaneity (Alderson-Day et al., 2018; Grandchamp et al., 2019; Pratts et al., 2023). Many of these varieties have been well-documented since Vygotsky (1934/1986), who provided foundational insights describing inner speech's abbreviated, dialogic, and socially derived forms as tools for self-regulation and cognitive mediation. However, while these phenomenological variations are well-established, theoretical frameworks have struggled to explain why, when, and how inner speech adopts specific configurations. By embedding these observations within a linguistic active inference

framework, LAIT goes beyond isolated descriptions to provide mechanistic explanations that unify inner speech's forms and functions through shared computational principles.

Specifically, LAIT proposes that these phenomenological varieties reflect functionally specialised computational strategies for active inference, with each variation - from condensed to expanded structures, monologic to dialogic exchanges, self to other voices, and spontaneous to deliberate forms - representing an optimised solution tailored to different inferential demands and contexts.

Condensed versus expanded structures

The variation between condensed and expanded inner speech reflects a *structural* distinction: providing unstructured categorical labels for perception and action, versus formulating structured causal and temporal sequences to orchestrate an unfolding cascade of active inference cycles.

Condensed inner speech uses abbreviated linguistic forms to act as high-precision priors to efficiently categorise sensory information and guide simple actions (Lupyan, 2012). Simple word labels are sufficient to facilitate categorical perception. For example, colour word labels help Russian and Mongolian speakers distinguish different shades of blue without requiring complex syntax (He et al., 2019; Winawer et al., 2007). In object recognition, hearing object names enhances early visual processing more effectively than nonverbal sounds (Rabovsky et al., 2012). Similarly, action words activate the motor cortex more than non-action words during word production (Oliveri et al., 2004). When participants heard action words like *"squeeze"* while observing the corresponding action, they directed more eye fixations to the target object and significantly increased motor evoked potential in relevant muscles, indicating enhanced attention control and action preparedness (Franklin et al., 2020). While direct observation of condensed inner speech remains challenging, these findings nevertheless suggest that minimal linguistic forms can sufficiently facilitate categorical perception and action guidance.

In contrast, expanded inner speech provides the syntactic and prosodic structures needed to articulate causal relationships and temporal sequences crucial for complex reasoning and

planning. At the syntactic level, the structure of expanded inner speech offers a framework for organising complex thought processes, specifying temporal order of events and causal relationships between elements (Gallo et al., 2009), particularly evident in problem-solving and planning tasks where participants report using full sentences to work through solutions (Morin et al., 2018). The syntactic structure works in concert with the rhythmic and intonational patterns of inner speech (Yao, 2025; Yao & Scheepers, 2015, 2018), which provide embodied predictive signals that guide the interpretation of complex sentence structure, support working memory maintenance, and potentially carry emotional meaning that enriches inference (Kreiner & Eviatar, 2024). These combined syntactic and prosodic features naturally extend into broader dialogic forms of inner speech, where multiple perspectives diverge and converge through dialectical reasoning.

Monologic versus dialogic forms

The distinction between monologic and dialogic inner speech comprises whether it unfolds in a single stream of verbal thought or involves multiple perspectives engaging in mental conversation (Fernyhough, 1996), which reflects two different approaches: *sequential* inference versus *reciprocal parallel* inference. Sequential inference means that perception-action cycles unfold one after another, with the posterior belief serving as the prior for the next cycle. Reciprocal parallel inference involves multiple concurrent streams of active inference cycles that interact dynamically through message passing, where each stream treats outputs (posteriors) from others as input evidence or constraints on its own generative model. This enables back-and-forth exchanges, facilitating the simultaneous consideration of perspectives.

Monologic inner speech maintains a single stream of thought for sequential inference, breaking down problems into a linear progression of sequential steps for structured uncertainty reduction. Although researchers have not yet established direct links between monologic inner speech and cognitive functions that are sufficiently supported by sequential inference, such as planning or sequential problem-solving (Baldo et al., 2005; Lidstone et al., 2010), its functional role can be inferred from research on dialogic inner speech.

Specifically, studies have identified task contexts that benefit particularly from dialogic inner speech, including abstract concept processing (Borghi & Fernyhough, 2022), social reasoning (Fernyhough, 2008), emotional regulation (Orvell et al., 2021), and self-reflection (Morin, 2018). These cognitive functions cannot be solved effectively through sequential inference alone because they inherently involve reflective, back-and-forth thinking that requires consideration of multiple perspectives simultaneously. The suggests, in turn, that tasks like planning or sequential problem-solving may be sufficiently supported by non-dialogic, i.e. monologic inner speech.

This differentiation is reflected in phenomenological surveys showing that monologic and dialogic inner speech are equally prevalent among the general population (Alderson-Day et al., 2018; McCarthy-Jones & Fernyhough, 2011). The prevalence of both forms in everyday experience implies that both are functionally important and are likely adaptively adopted according to varying task contexts and demands. This may explain the mixed evidence for dialogic inner speech in divergent thinking (de Rooij, 2022): tasks that only require generating multiple separate solutions can be accomplished through sequential inference, whereas tasks demanding simultaneous consideration of competing perspectives require the reciprocal parallel capacity of dialogic inner speech.

Self versus other voices

Inner speech flexibly recruits different voice characteristics, from self-voiced or un-voiced forms to simulations of others' voices, based on *perceptual cuing* demands for active inference.

The fundamental capabilities of inner speech - whether for sequential reasoning in monologic forms or parallel processing in dialogic forms - can operate effectively through self-voiced or un-voiced speech (Fernyhough, 2004). However, when additional perceptual cuing would benefit the inferential process, distinct voice characteristics may be recruited to help enrich and distinguish different streams of thought.

Evidence from descriptive experience sampling indicates that inner speech in others' voices is generally rare and is typically experienced as inner hearing (without motor action) during an inner dialogue; it often directly adopts another person's vocal characteristics or is emulated by

altering individuals' own voices to mimic those vocal qualities (Hurlburt et al., 2013). This phenomenological variation indicates that representing others' perspectives may require vocal differentiation to maintain distinct conceptual boundaries between self and other during internal dialogue.

This functional value of other-voiced inner speech becomes particularly apparent in social reasoning contexts. When simulating others' perspectives is crucial (Fernyhough, 2008), the generation of their distinctive voices provides rich perceptual cues that strengthen mental models through perceptual simulation. The simulation of others' voices activates associated personality traits, speaking styles, and likely responses, enriching perceptual inference through mirroring mechanisms (Frith & Frith, 2006). Neuroimaging evidence supports this specialised role, showing that processing others' direct speech specifically engages both auditory and ToM networks (Alderson-Day et al., 2020).

Further clues for the functional significance of other-voice recruitment come from voice-hearing and verbal hallucination research. Voice-hearing and verbal hallucinations typically involve others' voices and are interpreted as originating from external sources rather than the self (Woods et al., 2015). The consistent attribution of hallucinated voices to other entities suggests that voice characteristics serve as powerful cues for source attribution and perspective differentiation, even when the underlying neural mechanisms differ from typical inner speech (Alderson-Day & Fernyhough, 2015a; Brookwell et al., 2013).

Taken together, while voice characteristics aren't always necessary for dialogic thinking, they may serve as perceptual aids, providing distinct representational markers for different cognitive agents during inference.

Spontaneous versus deliberate generation

The variation across spontaneous and deliberate forms of inner speech reflects complementary aspects of *perceptually-driven* and *goal-directed* processes in linguistic active inference.

Spontaneous inner speech appears to emerge through *bottom-up*, current prediction error reduction in perceptual inference, where prediction errors trigger *automatic* linguistic predictions. These linguistic predictions can take the forms of simple verbal labels to more elaborate linguistic expressions, emerging rapidly and spontaneously in response to prediction errors. For example, unexpectedly seeing an ambiguous shape might trigger *"cat?"* as an automatic linguistic prediction to guide perceptual inference, or encountering an unexpected problem might prompt an automatic self-directed question *"what shall I do?"* to invite more deliberate forms of active control.

While direct evidence linking spontaneous inner speech to prediction errors remains scarce, this mechanism can be inferred from silent reading research. Studies show increased auditory cortex activation during silent reading of direct speech quotations compared to indirect speech (Yao et al., 2011; Yao & Scheepers, 2011). Written direct speech quotations (e.g., *She says, "I'm fine!"*) introduce uncertainty about the quoted speaker's mental state, as these utterance lack critical prosodic cues that convey pragmatic meaning, such as excitement, affirmation, or sarcasm (Clark & Gerrig, 1990). The brain fills this gap by generating top-down predictions of the missing prosodic cues, enabling readers to infer the quoted speakers' intended meaning and mental states (Yao et al., 2012). In contrast, when mental states are explicitly provided in quoted thoughts (e.g., *She thought…*), this uncertainty disappears, eliminating the need for perceptual simulation (Alderson-Day et al., 2020). These findings demonstrate how prediction errors, such as missing auditory cues in written direct speech, spontaneously trigger perceptual simulations of speech to aid inference.

Deliberate inner speech, in contrast, operates through *top-down*, prospective prediction error reduction in *goal-directed* active control, where we deliberately generate inner speech to meet task demands or achieve specific goals (Alderson-Day & Fernyhough, 2015a). Rather than responding to current perceptual mismatches, this process anticipates and minimises the predicted gap between present and desired future states. This goal-directedness is inherent in the vast majority of experimental paradigms where researchers either directly instruct participants to use inner speech (Scott, 2013; Tian et al., 2016) or design tasks with goals that naturally induce inner speech engagement. Such tasks can range from simple acts like counting or rehearsing a phone

number for memory retention (Baddeley, 2003) and differentiating visual stimuli based on their phonological properties (Geva & Warburton, 2019), to more complex tasks like working through problems or planning future scenarios with specific problem-solving objectives (Baldo et al., 2005; Lidstone et al., 2010). In each case, inner speech is deliberately generated to facilitate progress towards task-defined goals for active control.

This distinction between perceptually-driven and goal-directed processes does not imply constant and mutually exclusive forms of inner speech in a given task. Rather, inner speech can fluidly shift between these forms - from immediate reactions to current prediction errors, to deliberate simulations that anticipate future states. This dynamic involves iterative cycles where expected prediction errors guide action selection through outcome simulation, while actual observations continuously refine subsequent actions, creating an adaptive feedback loop as computational needs evolve.

From discrete to fluid forms of inner speech for dynamic computations

LAIT reveals how different forms of inner speech reflect specialised implementations of linguistic active inference, optimised for varying computational demands for prediction error and uncertainty reduction. The brain adaptively deploys different forms depending on the trigger (perceptually-driven vs. goal-directed prediction errors) and computational complexity: simpler forms (condensed, monologic, un-voiced) can emerge spontaneously to interpret current sensory inputs or are deliberately generated for straightforward goal-directed control, while more complex forms (expanded, dialogic, self/other-voiced) are constructed when resolving complex prediction errors or pursuing goals that require extended causal modelling, hierarchical structuring, or multi-perspective inference.

Rather than manifesting as discrete subtypes, LAIT highlights that inner speech exhibits fluid characteristics that are flexibly combined to address different inferential demands. Any instance may simultaneously engage multiple forms and characteristics - for example, combining expanded form for complex causal modelling with dialogic structure for parallel hypothesis testing, or mixing self and other voices to enhance social inference through embodied simulation. This

computational flexibility enables rapid adaptation to changing contexts and prediction errors across perceptual inference and active control.

These fluid inner speech forms dynamically shift as computational needs and prediction errors fluctuate. What begins as condensed speech during spontaneous perceptual inference might expand into dialogic form when inference fails or substantial prediction errors arise, which may recruit additional voice characteristics when social inference becomes relevant. This dynamic deployment of inner speech forms suggests moving beyond trait-based approaches to studying *real-time* state changes as computational requirements evolve.

Taken together, LAIT transforms phenomenological observations, moving beyond a descriptive catalogue of diverse forms to a unified framework that explains their underlying logic. That is, it posits that varied forms are specialised adaptations to distinct computational requirements, revealing systematic correspondences between phenomenological manifestations and inferential mechanisms. This explains why, when, and how inner speech can flexibly allocate cognitive resources across timescales (from immediate perceptual inference to long-term planning) and domains (from sensorimotor predictions to abstract reasoning). This mechanistic account provides a crucial advantage over theories framing inner speech as a general-purpose 'cognitive tool', as it generates clearer, falsifiable predictions and enables more targeted, theoretically-grounded investigations.

Unifying Diverse Models of Inner Speech

Beneath inner speech's diverse functions and phenomenology, LAIT reveals how neurocognitive and developmental models of inner speech could also be unified through the lens of linguistic active inference, representing different but complementary aspects of the same underlying predictive and inferential processes.

Neurocognitive Implementation of Linguistic Active Inference

Within LAIT, inner speech is a generative process that deploys linguistic predictions to transform priors and sensorimotor predictions for active inference. This process can manifest in conscious auditory experiences through two neurocognitive pathways - corollary discharge and

perceptual simulation. These auditory manifestations emerge from different computational roots: corollary discharge arises from the motor command for an intended articulation, providing auditory feedback during goal-directed active control; in contrast, perceptual simulation reactivates learned sound patterns from memory, broadcasting linguistic predictions that are deployed to resolve prediction errors during perceptually-driven inference. Both pathways facilitate conscious access to linguistic content, operating within working memory systems that maintain and manipulate linguistic representations for active inference.

Inner speech with corollary discharge. Corollary discharge refers to the predicted sensory outcome of an intended action (Sperry, 1950). Within the speech production system, corollary discharge provides the auditory content of motor-based inner speech (Scott, 2013). By generating an 'efference copy' of speech signals during covert speech production, the brain predicts its sensory outcomes, i.e. what it would sound like when spoken aloud, thus creating an internal phonological representation that we perceive as the sound of our 'inner speaking'.

The neural basis for this model is well-established through converging evidence from multiple methodologies. Functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) studies demonstrate activation in speech-production areas, particularly the L-IFG, during inner speech tasks (Aleman et al., 2005; Lurito et al., 2000; Shergill et al., 2001). Magnetoencephalography (MEG) studies show increased auditory activity ~170ms after imagined articulation, illustrating the temporal progression of motor-to-auditory transformations in the predictive pathway (Tian & Poeppel, 2010). The functional role of these predictions is demonstrated by Scott (2013), who showed that corollary discharge from inner speech can attenuate the perception of matching external sounds, indicating that inner speech shapes sensory processing through predictive mechanisms. Further supporting this, Jack et al. (2019) found that inner articulation of phonemes reduces the N1 ERP response to matching audible phonemes, indicating attenuated automatic sound processing, while mismatching phonemes do not produce this attenuation.

Within LAIT, inner speech that generates corollary discharges reflects how the brain deploys linguistic predictions for active control - providing the phonological feedback that monitors

our deliberate, goal-directed attempts to augment predictive processes for prediction error and uncertainty reduction. It is important to note that the audible speech experience provided by corollary discharge reflects rather than constitutes the underlying processes of active control. Active control involves broader predictive and inferential mechanisms beyond mere sounds, including goal formulation, action planning, outcome simulation, and self-regulation. These corollary discharges may provide perceivable representations for self-monitoring (*"What if I tried X?"*), thereby enabling self-evaluation and regulation (*"That wouldn't work because…"*) and enhancing conscious control over cycles of linguistic active inference.

Inner speech with perceptual simulation. Perceptual simulation provides a complementary mechanism to generate the auditory experience of perceptually-driven inner speech. The term 'simulation' here is borrowed from the grounded cognition literature, referring to the top-down reactivation of perceptual experiences from memory (Barsalou, 1999, 2008). In this context, perceptual simulation reactivates the audible sounds that accompany spontaneously generated inner speech – a language process that actively selects and deploys linguistic knowledge without formulating a motor plan (cf. De Livio et al., 2025). Unlike the deliberate, goal-directed predictions generated with corollary discharge for active control, this form of inner speech emerges spontaneously during perceptual inference to help interpret the causes of prediction errors, often manifesting as the experience of 'inner hearing'.

Neuroimaging studies reveal distinct neural signatures for this perceptual pathway of inner speech. FMRI findings from Tian et al. (2016) demonstrate a clear dissociation: while imagined speaking activates motor-to-perceptual transformation regions, imagined hearing specifically engages memory networks in the middle frontal and inferior parietal areas. This distinction is further supported by studies of silent reading. Yao et al. (2011) found that silent reading of direct speech quotes elicits vivid inner speech, resulting in increased activations in the auditory cortices. Yao et al. (2021) demonstrate that this reading-induced inner speech coincides with phase changes in theta-band oscillations in the auditory cortices, similar to those in actual speech perception. These findings collectively establish perceptual simulation as a complementary mechanism for generating inner speech sounds, without intentional motor involvement.

Within LAIT, inner speech generated with perceptually simulated sounds reflect spontaneous deployment of linguistic predictions for perceptual inference - providing phonological cues that 'broadcast' the detection of prediction errors to draw conscious attention, thereby inviting active control through deliberate inner speech. This interaction forms a cyclical active inference loop between perceptually-driven inner speech (supporting perceptual inference) and goal-directed inner speech (supporting active control).

Inner speech in working memory architecture. The phonological loop component of working memory (Baddeley, 2003; Baddeley & Hitch, 1974) provides the cognitive infrastructure for linguistic active inference, maintaining linguistic predictions and enabling their flexible manipulation and continuous interaction within a shared 'workspace'.

Inner speech reverberates within the phonological loop through the integrated operation of active manipulation and passive storage that synergise with its generation mechanisms. Active articulatory manipulation implements goal-directed inner speech, while the phonological store sustains both goal-directed and perceptually driven inner speech to enable their interface.

This functional organisation is reflected in the neural architecture: neuroimaging evidence reveals that articulatory rehearsal in working memory engages speech production areas including the L-IFG, premotor and supplementary motor areas, while the phonological store recruits regions specialised for phonological processing and speech representation in the left supramarginal gyrus (Paulesu et al., 1993; Smith & Jonides, 1998). This neural organisation overlaps with areas activated during corollary discharge and perceptual simulation (Pratts et al., 2023), indicating a shared neural infrastructure.

Working memory's role in supporting linguistic active inference is evidenced through two lines of research. On the one hand, disrupting working memory impacts linguistic active inference, as verbal interference impairs verbal and nonverbal reasoning (Baldo et al., 2005; Farmer et al., 1986; Phillips, 1999; Toms et al., 1993), perceptual categorisation (He et al., 2019; Winawer et al., 2007), and action control (Baddeley et al., 2001; Emerson & Miyake, 2003). On the other hand, research on individual differences reveals how working memory capacity constrains linguistic active inference capabilities - capacity predicts performance in predictive inference during reading

(Linderholm, 2002), classification and inference learning (Craig & Lewandowsky, 2013), and probabilistic inference about future events (Cashdollar et al., 2017). Neural evidence further supports this constraining role, with working memory load modulating activation patterns during inference processing (Virtue et al., 2008) and capacity-dependent theta oscillations during prospective uncertainty reduction (Cashdollar et al., 2017). Together, these findings demonstrate how working memory provides crucial infrastructure that enables and constrains linguistic active inference, with executive control orchestrating the dynamic allocation of these finite resources across cognitive operations (Carpenter et al., 2000; D'Esposito & Postle, 2015).

Implementing linguistic active inference. The integration of perceptual simulation and corollary discharge mechanisms within the working memory infrastructure illustrates how the brain implements the perceptual inference and active control aspects of linguistic active inference. Perceptual simulation signals spontaneous and often condensed inner speech for rapid perceptual inference. In contrast, corollary discharge reflects deliberate and often expanded inner speech for problem-solving and self-regulation.

These perceptually-driven and goal-directed modes of inner speech operate in continuous cycles, implementing the perception-action loops of active inference. As perceptually-driven inner speech detects and interprets prediction errors, it triggers goal-directed inner speech for active control. The planned actions generate expected sensory outcomes that guide action selection, while their eventual execution produces actual sensory feedback for the next inference cycle. This continuous interaction explains the fluid and dynamic transitions between different functions and forms of inner speech as computational demands shift.

Working memory provides the crucial computational workspace where these mechanisms interface and interact, with the phonological loop enabling sustained maintenance and manipulation of linguistic predictions through executive control. The coordinated operation of these neurocognitive systems creates an integrated linguistic active inference system whose emergence and evolution are further illuminated by Vygotsky's developmental theory.

Developmental Optimisation of Linguistic Active Inference

Vygotsky (1934/1987) established the social origins and developmental transformation of inner speech, which can be understood as a progressive optimisation of linguistic active inference, evolving from interpersonal social dialogue to increasingly efficient and flexible intrapersonal active inference.

This theoretical framework describes how prelinguistic intelligence transforms into a sophisticated symbolic system through the internalisation of social dialogue (Fernyhough, 2010; Luria, 1965). The resulting dialogic cognitive architecture provides the fundamental mechanism through which linguistic active inference operates - enabling prediction, reasoning, and self-regulation through cycles of internalised linguistic processes (Fernyhough & Borghi, 2023). Empirical support for this developmental trajectory comes from studies showing both dialogic and condensed forms of inner speech in adults (Alderson-Day et al., 2018; McCarthy-Jones & Fernyhough, 2011). These studies demonstrate inner speech's crucial role in executive functions and behavioural regulation (Cragg & Nation, 2010; Fernyhough, 2008), as well as in metacognition and self-awareness (Morin, 2005, 2022).

Social dialogue as interpersonal linguistic active inference. The foundation of linguistic active inference begins in social dialogue between child and caregiver, establishing an *interpersonal* perception-action loop. Through this dialogic framework, caregivers scaffold children's developing ability to use language for active inference.

This early phase features experience-driven learning of basic predictive frameworks through exposure to caregiver speech patterns and participation in joint actions and exchanges (Bruner, 1985; Saffran et al., 1996; Tomasello, 2005). For example, through verbal guidance and gesturing, caregivers guide children's attention to perceptual features (e.g., *"Look at the birds in the sky!"*, *"Look at their wings!"*), helping children to establish predictive models of what to attend to. Moreover, caregivers provide linguistic feedback to regulate the child's actions and emotions (e.g., *"Please don't bang your glass - it will break!"*, *"Are you hurt? It's okay - next time let's be more careful, OK?"*), helping children develop linguistic frameworks for understanding and regulating their behaviour and emotions.

This interpersonal dialogic exchange positions caregivers as external predictive controllers, using language to dynamically address prediction errors and guide the development of the child's predictive models. Research demonstrates that linguistic guidance and regulation increase during uncertain or challenging situations where prediction errors are likely to be highest (Lucca et al., 2019; Reuter et al., 2019). The sophistication of this linguistic scaffolding, as indicated by measures like vocabulary diversity, predicts children's later self-regulation abilities (Vallotton & Ayoub, 2011), highlighting its crucial role in developing predictive models for intrapersonal active inference.

Private speech emulates interpersonal linguistic active inference within the self. As children's language and cognitive capabilities develop, private speech, or overt self-directed talk, emerges as their attempt to emulate interpersonal linguistic active inference within the self. This transition marks a crucial shift from requiring external feedback loops involving caregivers to generating *intrapersonal* predictions and inferences through self-directed speech (Winsler, 2009). Drawing on their accumulated linguistic expertise, children begin deploying phrases and expressions previously used by caregivers to regulate themselves (Huttenlocher et al., 2010; Vallotton & Ayoub, 2011).

This developmental transition reveals early optimisation of linguistic active inference. Children's private speech initially mirrors elaborate dialogic patterns of social interaction. However, as their predictive models become more optimised, private speech gradually becomes more condensed and abbreviated, retaining key linguistic elements essential for active inference (Diaz et al., 1992; Winsler et al., 2009). This condensation demonstrates the progressive optimisation of linguistic active inference, refining predictive processing for greater efficiency while maintaining functional effectiveness.

Computational optimisation through internalisation. The internalisation of overt private speech to covert inner speech represents further optimisation of linguistic active inference, as children progress from overt self-regulation to silent linguistic mediation (Alderson-Day & Fernyhough, 2015a). Supported by developing inhibitory control (Kochanska et al., 1996),

internalisation eliminates the overhead of overt articulation and facilitates further condensation through abbreviated syntax and personalised semantics (Fernyhough & McCarthy-Jones, 2013).

The emergence of internalisation marks a key developmental stage in the optimisation of linguistic active inference. Studies show that developmentally at-risk preschool children exhibit less internalised private speech compared to typically developing peers, indicating delayed optimisation of linguistic active inference that still requires overt verbal support (Mulvihill et al., 2023). Children who successfully internalise their private speech demonstrate better self-regulation (Winsler et al., 2003), suggesting that internalisation improves both the efficiency and functionality of linguistic mediation. Individual differences in private speech internalisation (Winsler et al., 2009) could reflect varying trajectories in the optimisation of linguistic active inference, with the prevalence of condensed forms of inner speech in adults (McCarthy-Jones & Fernyhough, 2011) representing the mature stage of this process.

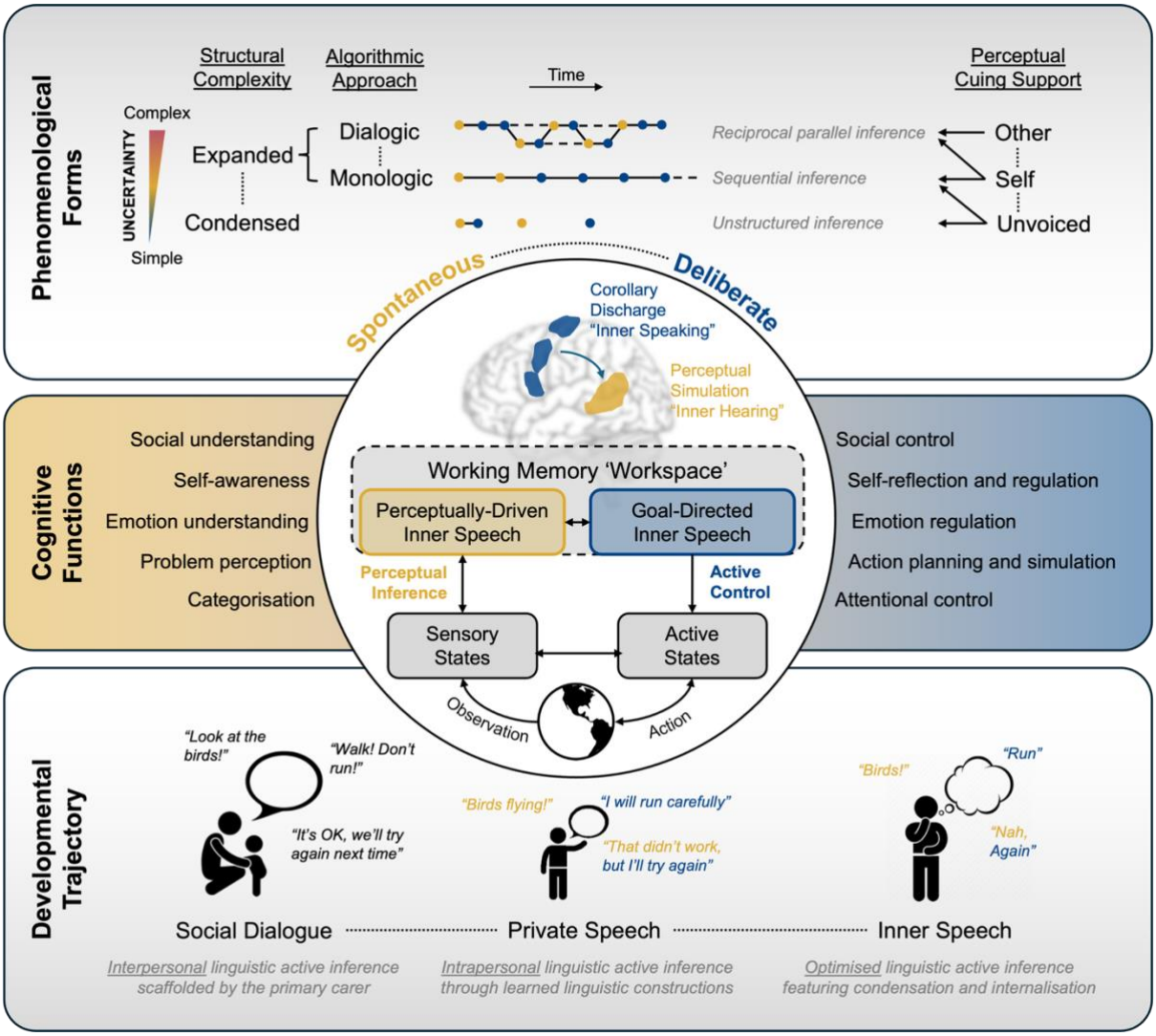
Whilst the Vygotskian framework describes a one-way developmental trajectory from private speech towards internalisation, adults with fully developed inner speech do revert to overt private speech, particularly when confronting complex or novel challenges. Adult private speech represents an adaptive deployment of linguistic active inference when computational demands favour the quality of linguistic predictions over their speed. Inner speech does not always activate the precise phonological representations of overt speech (Oppenheim & Dell, 2010), and its rapid, condensed nature can make it difficult to maintain and extend long inferential chains in memory. Overt private speech, in contrast, forces slower, more deliberate articulation that generates coherent linguistic formulations accompanied by high-fidelity auditory feedback. This external feedback loop creates a more persistent and precise memory trace that can be monitored and built upon across cycles of linguistic active inference. Adult private speech thus represents a dynamic trade-off, sacrificing the speed of internalisation for the stability and precision gained from engaging an external speech feedback loop – an optimal strategy when a task demands inferential chains that are persistent and extendable, rather than fleeting and fragmented.

Understanding Inner Speech as Linguistic Active Inference

In sum, LAIT reveals how inner speech's diverse cognitive functions, phenomenological forms, and neurocognitive-developmental mechanisms emerge coherently through linguistic active inference (see a visual synthesis in **Figure 2**). These functions emerge through inner speech's augmentation of perception-action cycles, with its phenomenological variations reflecting different computational requirements for inference and control, and the neurocognitive mechanisms reflect the dynamic deployment of linguistic predictions through perceptual and motor speech pathways supported by working memory, all operating within a coherent inferential framework. Rather than addressing diverse manifestations as separate phenomena with specialised explanations, LAIT unifies functional, phenomenological, and neurocognitive-developmental insights under shared computational principles and mechanisms, poised to deliver theoretical advances, generate novel testable hypotheses, motivate methodological innovations, and foster dialogues between previously isolated investigations.

Figure 2

Linguistic active inference: A unified explanation of inner speech's forms, functions, mechanisms and development.



Notes: The central circle depicts the perception-action loop of linguistic active inference supported by verbal working memory. Two complementary mechanisms generate inner speech: perceptual simulation ("inner hearing") supports perceptually-driven inner speech for perceptual inference (yellow pathway), while corollary discharge ("inner speaking") enables goal-directed processing for active control (blue pathway).

The top panel depicts phenomenological forms of inner speech varying along dimensions of condensation (condensed vs. expanded), diagicality (monologic vs. dialogic), voice qualities (unvoiced, self, other), and spontaneity (spontaneous, deliberate), reflecting different computational requirements for uncertainty reduction.

The middle panel shows examples of cognitive functions supported by linguistic active inference: perceptual inference functions (left, yellow) include categorisation, problem perception, socio-emotional and self-understanding, while active control functions (right, blue) include attentional control, action planning, and socio-emotional and self-regulation.

The bottom panel illustrates the developmental trajectory from interpersonal to intrapersonal linguistic active inference, showing how social dialogue scaffolded by caregivers evolves into private speech and eventually into optimised inner speech with increasing condensation and internalisation.

PART 3: Theoretical Advances and Research Implications

Building on the theoretical foundations of LAIT, this final section articulates its theoretical advances and translates them into a practical research roadmap. I will begin by detailing LAIT's capacity to unify and enhance current theoretical foundations, before formulating novel hypotheses about inner speech dynamics, outlining necessary methodological innovations to test them, and exploring the broader implications of this framework for related mental phenomena and symbolic thought.

Improving Explainability of Current Theoretical Foundations

The computational principles and mechanisms in LAIT improve the explainability of foundational theories of inner speech as a psychological tool for cognitive mediation (Fernyhough, 1996, 2010; Luria, 1965; Vygotsky, 1987). While influential, this 'cognitive tool' framework remains theoretically underspecified, lacking the principled mechanistic detail required for generating precise, falsifiable *a priori* hypotheses. The framework does not delineate which cognitive processes inner speech should enhance, articulate what 'mediation' entails mechanistically, or explain why such mediation is necessary - beyond the broad and difficult-to-falsify claim that it enhances cognition. This theoretical imprecision renders the framework overly accommodating, allowing researchers to posit inner speech mediation in virtually any cognitive task and to operationalise concepts like 'challenge' or 'mediation' with considerable latitude. Consequently, the framework struggles to generate specific predictions about why inner speech enhances some

cognitive functions but not others, or to reconcile the mixed findings on its relevance within domains such as categorisation (Fernyhough & Borghi, 2023).

Current frameworks observe that inner speech becomes more prevalent and expanded under cognitive challenges. Whilst the cognitive tool characterisation inherently predicts increased inner speech, it does not inherently predict that expanded forms should help with cognitive challenges. On the one hand, the concept of cognitive challenge remains vaguely defined and flexibly operationalised, often conflated with cognitive load and task complexity. On the other, the computational advantages of expanded versus condensed forms – whether they stem from phonological engagement or syntactic expansion, and how these differences should address which types of challenges - remain underspecified. This presumed link between expansion and cognitive challenges can also be dissociated. A highly challenging situation like fleeing a natural disaster might involve minimal, condensed inner speech, relying instead on sensorimotor intuition and impulses; conversely, a low-challenge scenario, such as debating whether an ambiguous shadow by the bins is a cat or a fox, might elicit expanded dialogic inner speech with little at stake.

LAIT advances the cognitive tool idea by specifying the computational drivers and mechanisms of inner speech, replacing the vague notions of ‘cognitive challenges’ and ‘cognitive mediation’ with a precise, operationalisable account of why and how inner speech augments cognition. The core driver is the brain’s imperative to reduce quantifiable imprecision (uncertainty) in its generative model - a more precise and operationalisable principle than ‘cognitive challenge’. This uncertainty reduction is implemented as inner speech modulates the content and precision of priors and their resulting sensorimotor predictions – a precise, quantifiable mechanism that provides the computational basis for ‘cognitive mediation’. This mechanism augments active inference in two important ways. One, it enhances efficiency by applying a compact linguistic label, thereby imposing a high-precision categorical prior, compressing a complex sensory input into a simplified, more tractable representation for inference. Two, it extends the scope of active inference by leveraging linguistic extendibility and generativity to build complex, and novel predictive models, orchestrating chains of inference for abstract and creative problem-solving beyond the limits of direct sensorimotor experience.

This computational specificity offers superior explanatory power. It explains *why* inner speech is engaged when model uncertainty is high and clarifies *how* it helps reduce that uncertainty. For instance, Russian speakers better discriminate subtle shades of blue through linguistically sharpened priors that produce more specific, distinct sensory predictions for finer differentiation (Winawer et al., 2007). Similarly, coupling novel actions with linguistic labels creates more distinct, precise predictive models for effective recall and reproduction (Gervasi et al., 2025). Conversely, this principle accounts for why linguistic augmentation provides little benefit when sensory input is unequivocally clear and model uncertainty is already low, as there is little imprecision to correct (Gerwien et al., 2022). This dynamic, uncertainty-driven engagement also helps explain discrepancies between trait-based self-report questionnaires and momentary descriptive experience sampling (Alderson-Day & Fernyhough, 2015b), as these methods capture different snapshots of an inherently fluctuating phenomenon.

Building from these core mechanisms, LAIT then provides a principled framework for predicting the contexts in which inner speech is beneficial and the forms it assumes, resulting from applying linguistic predictions to meet different inferential demands. For example, the framework predicts where inner speech is most relevant by linking the linguistic properties of efficiency, extendibility, and generativity with the computational advantages they confer in active inference. Inner speech is particularly useful for deploying efficient mental shortcuts to optimise cognitive resources and for constructing extended, creative mental models for abstract and novel problems beyond the reach of direct sensorimotor experience. In contrast, it is less critical for routine tasks that rely primarily on sensorimotor active inference, such as basic perceptual-motor coordination or familiar procedural activities.

Furthermore, the framework also specifies the computational requirements that determine its form. This clarifies, for instance, the presumed link between ‘challenge’ and ‘expansion’ with a specific, testable principle: inner speech expands when a task demands the linguistic formulation of causal and temporal structures for a complex predictive model. The principle explains, for instance, why reasoning through *“If I take this route, the traffic will be overwhelming at 5pm, so I*

should leave earlier" necessitates this expanded causal chain, whereas simple attentional control *"This way!"* does not.

Whilst LAIT improves the explainability of the cognitive tool account, both frameworks can appear to converge in their predictions when tested through crude experimental manipulations like verbal interference. This apparent overlap should not be mistaken for theoretical redundancy; rather, it highlights the limitations of a framework that too readily accommodates diverse outcomes through *post hoc* theoretical adjustments. The distinctive contribution of LAIT, therefore, lies in its capacity to generate specific, falsifiable *a priori* hypotheses and to motivate a more targeted research programme that systematically investigates uncertainty, tracks prediction errors, and examines inner speech manifestations in relation to their underlying computational mechanisms. This computational grounding enables precise insights into the nature, mechanisms, and dynamic manifestations of inner speech with a level of precision that the broader cognitive tool framework does not inherently guide researchers towards.

Generating Novel Testable Hypotheses

LAIT characterises inner speech as a dynamic, context-sensitive state that adapts to fluctuating uncertainty. This framing not only reveals the existing approaches are insufficient to capture inner speech's dynamic nature, but also generates novel, testable hypotheses rooted in the computational principles driving inner speech and the mechanisms governing its manifestations.

Uncertainty-Reducing Dynamic State Hypothesis. Inner speech is driven by the brain's need to reduce uncertainty - quantifiable imprecision in its generative model. This hypothesis generates the testable prediction that inner speech occurrence will fluctuate with model uncertainty, particularly in abstract domains requiring displaced and novel causal modelling. An individual's inner speech use will vary across different computational contexts, increasing in high-uncertainty situations - whether during tasks involving abstract reasoning, prospection, unfamiliar problems, or during specific task epochs, such as trials with unpredictable stimuli, or initial uncertain phases of problem-solving. Across individuals, those reporting limited access to inner

speech will exhibit more pronounced performance decrements specifically under high-uncertainty conditions, with minimal differences during routine sensorimotor processes. This hypothesis would be falsified if these predicted patterns are not observed.

Form-Computation Correspondence Hypothesis. Phenomenological variations of inner speech are determined by their underlying computational needs. That is, the need for causal and temporal structures recruits expanded inner speech; the need for parallel inference recruits dialogic forms; and the need for perceptual cuing recruits voice and prosodic variations. This would be falsified if experimentally manipulated computational demands (e.g., for causal structuring) fail to elicit the predicted changes in inner speech forms.

Perception-Action Cycling Hypothesis. Inner speech manifestations evolve across iterative cycles of perceptual inference and active control, implemented through dynamic engagement of perceptual and motor speech circuits. This predicts that inner speech will show systematic temporal patterns, transitioning between spontaneous, condensed forms during error-driven perceptual inference to deliberate, expanded forms for causal modelling and action selection and planning. Familiar situations will elicit rapid cycles, while novel situations will necessitate slower, more expanded cycles. This would be falsified if these systematic temporal patterns fail to manifest, or if the predicted shifts in neural engagement are not observed.

In sum, these three hypotheses establish inner speech as a measurable, context-sensitive dynamic process. Unlike descriptive accounts, LAIT anchors its hypotheses in precise computational principles and mechanisms. It systematically predicts when different forms emerge, why they shift dynamically, and how they relate to specific cognitive functions and neural engagement, creating clear opportunities for empirical validation and falsification.

Methodological Transformation and Collaborative Integration

Inner speech research currently suffers from methodological fragmentation rooted in theoretical underspecification, leading to isolated findings from phenomenological, functional, and neurocognitive studies that are difficult to integrate coherently. LAIT addresses this by establishing linguistic active inference as a shared computational foundation, reframing these disparate

observations as manifestations of the same underlying process: linguistic augmentation of active inference to reduce uncertainty. This unified framework motivates a methodological transformation, shifting research from descriptive cataloguing to mechanistic investigation.

This transformation requires moving beyond static, trait-based approaches to instead capture inner speech as a dynamic, context-sensitive process. For instance, phenomenological research can transition from survey-based documentation to theory-driven dynamic process tracking, predicting when and why specific forms - from condensed monitoring (*"Traffic ahead"*) to expanded reasoning (*"If I take the bypass…"*) - should emerge as computational demands and uncertainty shift. Similarly, functional studies can move beyond crude verbal interference paradigms to manipulate and measure the specific computational drivers of inner speech. Instead of simply blocking language, researchers can use strategic prompting (e.g., *"What's the alternative?"*) to elicit specific forms like dialogic inner speech and track how this impacts performance. Concurrently, neuroscientific research can advance beyond artificial, reductive tasks to map how perceptual and motor speech circuits engage during naturalistic cognition, providing objective neural signatures of these dynamic computational processes.

LAIT's greatest value lies in motivating interdisciplinary, multi-method integration that triangulate the nature and mechanism of inner speech. Consider planning a dinner party for guests with conflicting dietary needs. An integrated approach would combine methods to reveal a coherent computational narrative: phenomenological reports would show inner speech transitioning from condensed notes (*"Rice?"*, *"Donuts?"*) to expanded, dialogic reasoning; linguistic analysis would identify a corresponding increase in conditional and perspective-taking language; and neuroimaging would reveal dynamic recruitment of speech, theory-of-mind, and executive control networks. Crucially, performance metrics would then validate whether these coherent shifts improve inference efficiency and solution quality, linking form, content, and neural activity to a clear computational benefit.

To realise this integrative approach, the field requires significant methodological innovation. First, we need dynamic state tracking methods, like experience sampling timed to experimental manipulations, to capture inner speech fluctuations in theoretically guided ways. Second, we must

also develop paradigms to systematically manipulate and measure uncertainty and track resulting prediction errors through neurophysiological markers or behavioural indicators. Third, naturalistic neural mapping using high-temporal resolution techniques (EEG/MEG) is needed to connect dynamic neural network reconfigurations to specific linguistic and cognitive operations during ecologically valid tasks. Such innovations, driven by strategic cross-disciplinary collaboration, can transform our ability to study the dynamic relationship between language, thought, and adaptive behaviour.

Generalisation and Limitations Across Symbolic Systems

While LAIT was developed to explain inner speech, its core principle - that symbolic systems modulate predictive models to reduce uncertainty - potentially generalises beyond spoken language. That is, any sufficiently developed symbolic system could support similar processes if it provides a generative grammar for modelling the world and for mapping symbols from abstract thoughts to sensorimotor experiences. For instance, congenitally deaf individuals likely employ inner sign language or an abstract form of ‘speech’ for active inference in ways parallel to hearing individuals’ use of inner speech (Zimmermann & Brugger, 2013). Similarly, individuals who report predominantly visual thinking (Nedergaard & Lupyan, 2024) may use what Barsalou terms ‘perceptual symbols’ to implement active inference (Barsalou, 1999). This flexibility in symbolic active inference is evident across domains: Mathematicians appear to favour geometric shapes for visual reasoning before encoding them to symbols (Noss & Hoyles, 1995), and can flexibly use verbal and mathematical symbolic systems to solve equivalent problems (Sohn et al., 2004). Musicians similarly develop their own symbolic representations, using or even inventing notations to think about music (Barrett, 2004) and evaluate them to derive meaning (Hultberg, 2002).

Importantly, this generalised ‘symbolic active inference’ perspective does not imply that all symbolic systems operate identically. Different representational formats and organisations likely confer distinct computational properties that shape active inference. While verbal symbols excel at sequential processing and categorical abstraction, visual symbols leverage parallel processing and spatial relationships. Mathematical symbols may offer precise quantitative representation but lack

the intuitive accessibility of visual forms. Musical notation may uniquely capture temporal-acoustic patterns while being less suited for other sensorimotor domains.

Moreover, it is crucial to note that symbolic active inference, while computationally efficient in many contexts, is not always optimal or even helpful. In situations demanding immediate, fine-grained sensorimotor predictions, symbolic representations can introduce an unnecessary layer of abstraction that compromises both the speed and quality of the predictions essential for peak performance. In martial arts, dance, or rock climbing, for instance, attempting to augment experience through language or other symbolic representations could disrupt the rapid, detailed sensorimotor predictions needed for such activities. While a martial artist might benefit from symbolic reasoning during training or strategic planning, in the immediate context of sparring, linguistic or symbolic mediation could impede the rapid, nuanced, and instinctive physical movements required.

This perspective has important implications for understanding individual differences in active inference. The effectiveness of any particular system likely depends on both individual expertise and task demands. For instance, while articulatory suppression may disrupt inner speech, individuals could switch to alternative symbolic or analogic systems, such as sign language or visual imagery, depending on their availability and efficacy. The impact of articulatory interference would thus depend on factors such as the relative efficiency of these alternative systems for the specific task and the individual proficiency in using them (Nedergaard & Lupyan, 2024). In other words, research on cognitive processes should consider the potential for multiple abstract systems to support active inference, rather than assuming the primacy of any single system. This opens important questions about how various symbolic and analogic systems might complement or compete to support cognition, and how individual proficiency in employing these systems for active inference might predict cognitive performance across different domains.

Potential Applications to Phenomena Related to Inner Speech

LAIT may offer new insights into other mental phenomena related to inner speech. Take verbal hallucinations for instance - while existing models focus primarily on explaining how self-

generated inner speech is misattributed to an external source (Frith, 1992), LAIT suggests this misattribution leads to disrupted perception-action cycles that are unable to reduce prediction errors. It is well established that verbal hallucinations can emerge from a breakdown in differentiating internally generated predictions (corollary discharge) from external sensory inputs (external input). In linguistic active inference, this breakdown creates a self-reinforcing cycle where an initial corollary discharge is misperceived as an external input, triggering a new active inference cycle to interpret this unexpected 'external' input. The resulting linguistic predictions, meant to resolve the mis-perceived input, are once again misperceived, creating a recursive cycle in which linguistic predictions are generated to address prediction errors that they themselves have caused. As such, verbal hallucinations might represent a scenario where linguistic active inference becomes trapped in a recursive loop of active control, potentially explaining both their persistent nature and often uncertainty-related content.

Similar explanations might illuminate repetitive negative thinking patterns in verbal rumination. From LAIT's perspective, rumination could reflect linguistic active inference becoming paralysed in negative perceptual inference, repeatedly categorising and interpreting threats without progressing to actionable plans (active control). This creates a self-perpetuating cycle where each attempt at linguistic perceptual inference results in new prediction errors rather than resolving existing ones. The system becomes trapped in a recursive loop of threat detection and elaboration, particularly when facing uncertainties that resist practical resolution, such as existential concerns or situations beyond individual control.

LAIT may also provide new insights into verbal mind wandering. Rather than viewing it solely as an attention control failure, we can interpret it as the mind prioritising the resolution of ongoing, more pressing uncertainties over immediate task demands. When immediate tasks pose minimal uncertainty or challenge, the mind may redirect focus to address prediction error signals from larger background uncertainties, such as personal concerns or future planning. This shift would manifest as task-unrelated inner speech attempting to identify and resolve these uncertainties.

The above examples illustrate a few instances where LAIT's principles might offer new perspectives on inner speech-related phenomena. While a thorough application of LAIT to these phenomena lies beyond the scope of this paper, these conjectures nevertheless outline promising directions for future research. Future studies might profitably explore how individual differences in linguistic active inference relate to susceptibility to these phenomena, and how understanding their relationship to uncertainty processing might inform intervention strategies.

Conclusion

LAIT redefines the study of inner speech by specifying its computational architecture, moving beyond phenomenological description to reveal its underlying principles and mechanisms. At its core, the framework proposes that inner speech augments the brain's predictive processes by linguistically transforming prior expectations to steer the cycles of perception and action that resolve prediction errors. Language's unique properties - its efficiency in encoding complex experiences, its extendibility across time and space, and its generativity in constructing novel predictions – then vastly extend the scope and capabilities of this process. The result is a unified, multi-level account that connects the computational imperative to reduce uncertainty to its algorithmic operation in cycles of perceptual inference and active control, and its neurophysiological implementation within the neural dynamics of speech and working memory.

The framework's primary contribution lies in its explanatory and predictive power: it not only unifies disparate findings but also generates a rich set of novel, falsifiable hypotheses regarding the dynamics, form, and function of inner speech. First, it characterises inner speech as a context-dependent dynamic state driven by the need to reduce uncertainty. Second, it proposes that phenomenological forms are recruited to meet specific computational needs. Third, it specifies how inner speech evolves across iterative cycles of perceptual inference and active control, with systematic temporal patterns transitioning across form, function and neural engagement.

Theoretically, these hypotheses move the field beyond descriptive catalogues by providing the precise principles and mechanisms that determine when different forms emerge, why they shift dynamically, and how they relate to specific cognitive functions and neural engagement, creating

clear opportunities for empirical validation and falsification. Methodologically, testing these predictions motivates a shift from static, trait-based approaches to dynamic methods that capture context-sensitive fluctuations; from crude verbal interference to paradigms that manipulate uncertainty and computational demands; and from reductive tasks to mapping neural dynamics during naturalistic cognition. The goal, therefore, is to establish an integrated, interdisciplinary scientific enterprise that triangulates how inner speech form, content, function, and neural dynamics combine to produce a clear computational benefit.

The stakes of such an enterprise are not just theoretical. In clinical settings, LAIT could offer valuable practical implications by reframing conditions like verbal hallucinations and rumination as maladaptive cycles of linguistic active inference that fail to resolve prediction errors. More broadly, it provides a conceptual sketch for a generalised theory of symbolic thought, suggesting how other symbolic systems might similarly augment active inference. This framework thus positions inner speech not as a mere cognitive curiosity, but as a prime exemplar for investigating the fundamental relationships between symbolic thought, adaptive behaviour, and human cognition.

CRediT statement

BY: Conceptualisation, Writing-Original draft preparation, Writing-Reviewing and Editing, Visualisation, Funding acquisition.

References

Abdel Rahman, R., & Sommer, W. (2008). Seeing what we know and understand: How knowledge shapes perception. *Psychonomic Bulletin and Review*, *15*(6), 1055–1063. Scopus. https://doi.org/10.3758/PBR.15.6.1055

Adams, R. A., Shipp, S., & Friston, K. J. (2013). Predictions not commands: Active inference in the motor system. *Brain Structure and Function*, *218*(3), 611–643. https://doi.org/10.1007/s00429-012-0475-5

Alderson-Day, B., & Fernyhough, C. (2015a). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin*, *141*(5), 931–965. https://doi.org/10.1037/bul0000021

Alderson-Day, B., & Fernyhough, C. (2015b). Relations among questionnaire and experience sampling measures of inner speech: A smartphone app study. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/article/10.3389/fpsyg.2015.00517

Alderson-Day, B., Mitrenga, K., Wilkinson, S., McCarthy-Jones, S., & Fernyhough, C. (2018). The varieties of inner speech questionnaire – Revised (VISQ-R): Replicating and refining links between inner speech and psychopathology. *Consciousness and Cognition*, *65*, 48–58. https://doi.org/10.1016/j.concog.2018.07.001

Alderson-Day, B., Moffatt, J., Bernini, M., Mitrenga, K., Yao, B., & Fernyhough, C. (2020). Processing speech and thoughts during silent reading: Direct reference effects for speech by fictional characters in voice-selective auditory cortex and a Theory-of-Mind network. *Journal of Cognitive Neuroscience*, *32*(9), 1637–1653. https://doi.org/10.1162/jocn_a_01571

Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Haan, E. H. F., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imagined spoken words. *Cerebral Cortex*, *15*(2), 221–228. https://doi.org/10.1093/cercor/bhh124

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology*, *35*(5), 1311–1320. https://doi.org/10.1037//0012-1649.35.5.1311

Baars, B. J. (1997). In the theatre of consciousness. Global Workspace Theory, a rigorous scientific theory of consciousness. *Journal of Consciousness Studies*, *4*(4), 292–309.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, *130*(4), 641.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baldo, J. V., Dronkers, N. F., Wilkins, D., Ludy, C., Raskin, P., & Kim, J. (2005). Is problem solving dependent on language? *Brain and Language*, *92*(3), 240–250. https://doi.org/10.1016/j.bandl.2004.06.103

Baldo, J. V., Paulraj, S. R., Curran, B. C., & Dronkers, N. F. (2015). Impaired reasoning and problem-solving in individuals with language impairment due to aphasia or language delay. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01523

Banks, B., & Connell, L. (2024). Access to inner language enhances memory for events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Bar, M., Tootell, R. B. H., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., Rosen, B. R., & Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*, *29*(2), 529–535. https://doi.org/10.1016/S0896-6273(01)00224-0

Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, *12*(1), 1–23. https://doi.org/10.1093/scan/nsw154

Barrett, M. S. (2004). Thinking about the representation of music: A case-study of invented notation. *Bulletin of the Council for Research in Music Education*, *161/162*, 19–28.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1281–1289. https://doi.org/10.1098/rstb.2008.0319

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*.

Benrimoh, D., Parr, T., Vincent, P., Adams, R. A., & Friston, K. (2018). Active inference and auditory hallucinations. *Computational Psychiatry (Cambridge, Mass.)*, *2*, 183.

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2019). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews*, *29*, 120–153. https://doi.org/10.1016/j.plrev.2018.12.001

Borghi, A. M., & Fernyhough, C. (2022). Concepts, abstractness and inner speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *378*(1870), 20210371. https://doi.org/10.1098/rstb.2021.0371

Boroditsky, L. (2018). Language and the construction of time through space. *Trends in Neurosciences*, *41*(10), 651–653. https://doi.org/10.1016/j.tins.2018.08.004

Boyd, R., Richerson, P. J., & Henrich, J. (2011). The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences*, *108*(supplement_2), 10918–10925. https://doi.org/10.1073/pnas.1100290108

Brookwell, M. L., Bentall, R. P., & Varese, F. (2013). Externalizing biases and hallucinations in source-monitoring, self-monitoring and signal detection studies: A meta-analytic review. *Psychological Medicine*, *43*(12), 2465–2475. https://doi.org/10.1017/S0033291712002760

Bruner, J. (1985). Child's talk: Learning to use language. *Child Language Teaching and Therapy*, *1*(1), 111–114.

Carpenter, P. A., Just, M. A., & Reichle, E. D. (2000). Working memory and executive function: Evidence from neuroimaging. *Current Opinion in Neurobiology*, *10*(2), 195–199. https://doi.org/10.1016/S0959-4388(00)00074-X

Carruthers, P. (2018). *The causes and contents of inner speech* (Vol. 1). Oxford University Press. https://doi.org/10.1093/oso/9780198796640.003.0002

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*(2), 579–593. https://doi.org/10.1016/j.cognition.2007.03.004

Casasanto, D., & Chrysikou, E. G. (2011). When left is "right": Motor fluency shapes abstract concepts. *Psychological Science*, *22*(4), 419–422. https://doi.org/10.1177/0956797611401755

Cashdollar, N., Ruhnau, P., Weisz, N., & Hasson, U. (2017). The role of working memory in the probabilistic inference of future sensory events. *Cerebral Cortex*, *27*(5), 2955–2969. https://doi.org/10.1093/cercor/bhw138

Chomsky, N. (1965). *Aspects of the theory of syntax* (Issue 11). MIT press.

Cibelli, E., Xu, Y., Austerweil, J. L., Griffiths, T. L., & Regier, T. (2016). The Sapir-Whorf hypothesis and probabilistic inference: Evidence from the domain of color. *PLOS ONE*, *11*(7), e0158725. https://doi.org/10.1371/journal.pone.0158725

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, H. H., & Gerrig, R. J. (1990). Quotations as Demonstrations. *Language*, *66*(4), 764. https://doi.org/10.2307/414729

Connell, L. (2019). What have labels ever done for us? The linguistic shortcut in conceptual processing. *Language, Cognition and Neuroscience*, *34*(10), 1308–1318.

Constant, A., Ramstead, M. J., Veissière, S. P., & Friston, K. (2019). Regimes of expectations: An active inference model of social conformity and human decision making. *Frontiers in Psychology*, *10*, 679.

Cragg, L., & Nation, K. (2010). Language and the development of cognitive control. *Topics in Cognitive Science*, *2*(4), 631–642. https://doi.org/10.1111/j.1756-8765.2009.01080.x

Craig, S., & Lewandowsky, S. (2013). Working memory supports inference learning just like classification learning. *Quarterly Journal of Experimental Psychology*, *66*(8), 1493–1503. https://doi.org/10.1080/17470218.2013.818703

Crawford, L. E. (2009). Conceptual metaphors of affect. *Emotion Review*, *1*(2), 129–139. https://doi.org/10.1177/1754073908100438

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148–153. https://doi.org/10.1016/j.tics.2009.01.005

D'Argembeau, A., Renaud, O., & Van der Linden, M. (2011). Frequency, characteristics and functions of future-oriented thoughts in daily life. *Applied Cognitive Psychology*, *25*(1), 96–103. https://doi.org/10.1002/acp.1647

De Livio, C., Borghi, A. M., & Fernyhough, C. (2025). Inner speech is not a simulation of language but an *act* of speaking: Comment on "The Sound of Thought: Form Matters – The Prosody of Inner Speech" by Hamutal Kreiner, Zohar Eviatar. *Physics of Life Reviews*, *53*, 218–220. https://doi.org/10.1016/j.plrev.2025.03.013

de Rooij, A. (2022). Varieties of inner speech and creative potential. *Imagination, Cognition and Personality*, *41*(4), 460–489. https://doi.org/10.1177/02762366211070999

Dehaene, S., Charles, L., King, J.-R., & Marti, S. (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology*, *25*, 76–84. https://doi.org/10.1016/j.conb.2013.12.005

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*(Volume 66, 2015), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031

Diaz, R. M., Berk, L. E., & Diaz, R. (1992). *Private speech: From social interaction to self-regulation*. L. Erlbaum Hillsdale, NJ.

Dove, G. (2018). Language as a disruptive technology: Abstract concepts, embodiment and the flexible mind. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1752), 20170135. https://doi.org/10.1098/rstb.2017.0135

Dove, G., Barca, L., Tummolini, L., & Borghi, A. M. (2022). Words have a weight: Language as a source of inner grounding and flexibility in abstract concepts. *Psychological Research*, *86*(8), 2451–2467. https://doi.org/10.1007/s00426-020-01438-6

Emerson, M. J., & Miyake, A. (2003). The role of inner speech in task switching: A dual-task investigation. *Journal of Memory and Language*, *48*(1), 148–168.

Estes, Z., & Ward, T. B. (2002). The emergence of novel attributes in concept modification. *Creativity Research Journal*, *14*(2), 149–156. https://doi.org/10.1207/S15326934CRJ1402_2

Farmer, E. W., Berman, J. V. F., & Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratch-pad in working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(4), 675–688. https://doi.org/10.1080/14640748608401620

Ferguson, B., & Waxman, S. (2017). Linking language and categorization in infancy. *Journal of Child Language*, *44*(3), 527–552. https://doi.org/10.1017/S0305000916000568

Fernyhough, C. (1996). The dialogic mind: A dialogic approach to the higher mental functions. *New Ideas in Psychology*, *14*(1), 47–62. https://doi.org/10.1016/0732-118X(95)00024-B

Fernyhough, C. (2004). Alien voices and inner dialogue: Towards a developmental account of auditory verbal hallucinations. *New Ideas in Psychology*, *22*(1), 49–68. https://doi.org/10.1016/j.newideapsych.2004.09.001

Fernyhough, C. (2008). Getting Vygotskian about theory of mind: Mediation, dialogue, and the development of social understanding. *Developmental Review*, *28*(2), 225–262. https://doi.org/10.1016/j.dr.2007.03.001

Fernyhough, C. (2010). Vygotsky, Luria, and the social brain. In B. Sokol, U. Muller, J. Carpendale, A. Young, & G. Iarocci (Eds), *Self- and Social-Regulation: Exploring the Relations Between Social Interaction, Social Understanding, and the Development of Executive Functions* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195327694.003.0003

Fernyhough, C. (2016). *The voices within: The history and science of how we talk to ourselves*. Basic Books.

Fernyhough, C., & Borghi, A. M. (2023). Inner speech as language process and cognitive tool. *Trends in Cognitive Sciences*, *27*(12), 1180–1193. https://doi.org/10.1016/j.tics.2023.08.014

Fernyhough, C., & McCarthy-Jones, S. (2013). Thinking aloud about mental voices. In F. Macpherson & D. Platchias (Eds), *Hallucination: Philosophy and Psychology* (p. 0). The MIT Press. https://doi.org/10.7551/mitpress/9780262019200.003.0005

Fini, C., Zannino, G. D., Orsoni, M., Carlesimo, G. A., Benassi, M., & Borghi, A. M. (2022). Articulatory suppression delays processing of abstract words: The role of inner speech. *Quarterly Journal of Experimental Psychology*, *75*(7), 1343–1354. https://doi.org/10.1177/17470218211053623

Franklin, Z. C., Wright, D. J., & Holmes, P. S. (2020). Using Action-congruent Language Facilitates the Motor Response during Action Observation: A Combined Transcranial Magnetic Stimulation and Eye-tracking Study. *Journal of Cognitive Neuroscience*, *32*(4), 634–645. https://doi.org/10.1162/jocn_a_01510

Friston, K. (2008). Hierarchical Models in the Brain. *PLOS Computational Biology*, *4*(11), e1000211. https://doi.org/10.1371/journal.pcbi.1000211

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), Article 2. https://doi.org/10.1038/nrn2787

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O Doherty, J., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, *68*, 862–879. https://doi.org/10.1016/j.neubiorev.2016.06.022

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, *29*(1), 1–49. https://doi.org/10.1162/NECO_a_00912

Frith, C. D. (1992). *The cognitive neuropsychology of schizophrenia.*

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531–534. https://doi.org/10.1016/j.neuron.2006.05.001

Gallo, I. S., Keil, A., McCulloch, K. C., Rockstroh, B., & Gollwitzer, P. M. (2009). Strategic automation of emotion regulation. *Journal of Personality and Social Psychology*, *96*(1), 11–31. https://doi.org/10.1037/a0013460

Gelman, S. A., & Meyer, M. (2011). Child categorization. *WIREs Cognitive Science*, *2*(1), 95–105. https://doi.org/10.1002/wcs.96

Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind: Advances in the study of language and thought*. The MIT Press. https://doi.org/10.7551/mitpress/4117.001.0001

Gentner, D., Imai, M., & Boroditsky, L. (2002). As time goes by: Evidence for two systems in processing space → time metaphors. *Language and Cognitive Processes*, *17*(5), 537–565. https://doi.org/10.1080/01690960143000317

Gervasi, A. M., Mazzuca, C., Borghi, A., & Brozzoli, C. (2025). Interfering with inner speech during action encoding impacts their execution. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *47*(0). https://escholarship.org/uc/item/13v0442r

Gervits, F., Johanson, M., & Papafragou, A. (2023). Relevance and the role of labels in categorization. *Cognitive Science*, *47*(12), e13395. https://doi.org/10.1111/cogs.13395

Gerwien, J., von Stutterheim, C., & Rummel, J. (2022). What is the interference in "verbal interference"? *Acta Psychologica*, *230*, 103774. https://doi.org/10.1016/j.actpsy.2022.103774

Geva, S., & Warburton, E. A. (2019). A Test Battery for Inner Speech Functions. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *34*(1), 97–113. https://doi.org/10.1093/arclin/acy018

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, *23*(5), 389–407. https://doi.org/10.1016/j.tics.2019.02.003

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565. https://doi.org/10.3758/BF03196313

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *WIREs Cognitive Science*, *1*(1), 69–78. https://doi.org/10.1002/wcs.26

Gordon, P. (2004). Numerical cognition without words: Evidence from amazonia. *Science*, *306*(5695), 496–499. https://doi.org/10.1126/science.1094492

Granato, G., Borghi, A. M., & Baldassarre, G. (2020). A computational model of language functions in flexible goal-directed behaviour. *Scientific Reports*, *10*(1), 1–13. https://doi.org/10.1038/s41598-020-78252-y

Granato, G., Borghi, A. M., Mattera, A., & Baldassarre, G. (2022). A computational model of inner speech supporting flexible goal-directed behaviour in Autism. *Scientific Reports*, *12*(1). Scopus. https://doi.org/10.1038/s41598-022-18445-9

Grandchamp, R., Rapin, L., Perrone-Bertolotti, M., Pichat, C., Haldin, C., Cousin, E., Lachaux, J.-P., Dohen, M., Perrier, P., Garnier, M., Baciu, M., & Lœvenbruck, H. (2019). The ConDialInt Model: Condensation, Dialogality, and Intentionality Dimensions of Inner Speech Within a Hierarchical Predictive Control Framework. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02019

Granito, C., Scorolli, C., & Borghi, A. M. (2015). Naming a lego world. The role of language in the acquisition of abstract concepts. *PLOS ONE*, *10*(1), e0114615. https://doi.org/10.1371/journal.pone.0114615

Hale, C. M., & Tager-Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science*, *6*(3), 346–359. https://doi.org/10.1111/1467-7687.00289

Harnad, S. (1987). *Psychophysical and cognitive aspects of categorical perception: A critical overview* (S. Harnad, Ed.; pp. 1–25). Cambridge University Press. https://eprints.soton.ac.uk/250386/

He, H., Li, J., Xiao, Q., Jiang, S., Yang, Y., & Zhi, S. (2019). Language and color perception: Evidence from Mongolian and Chinese speakers. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00551

Hockett, C. F. (1960). The origin of speech. *Scientific American*, *203*(3), 88–97.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*(3), 495–514. https://doi.org/10.3758/PBR.15.3.495

Hu, H., Yarats, D., Gong, Q., Tian, Y., & Lewis, M. (2019). Hierarchical decision making by generating and following natural language instructions. *Advances in Neural Information Processing Systems*, *32*. https://proceedings.neurips.cc/paper/2019/hash/7967cc8e3ab559e68cc944c44b1cf3e8-Abstract.html

Huang, W., Xia, F., Xiao, T., Chan, H., Liang, J., Florence, P., Zeng, A., Tompson, J., Mordatch, I., Chebotar, Y., Sermanet, P., Brown, N., Jackson, T., Luu, L., Levine, S., Hausman, K., &

Ichter, B. (2022). *Inner monologue: Embodied reasoning through planning with language models* (No. arXiv:2207.05608). arXiv. https://doi.org/10.48550/arXiv.2207.05608

Hultberg, C. (2002). Approaches to music notation: The printed score as a mediator of meaning in Western tonal tradition. *Music Education Research*, *4*(2), 185–197. https://doi.org/10.1080/1461380022000011902

Hurlburt, R. T., Heavey, C. L., & Kelsey, J. M. (2013). Toward a phenomenology of inner speaking. *Consciousness and Cognition*, *22*(4), 1477–1494. https://doi.org/10.1016/j.concog.2013.10.003

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Jack, B. N., Le Pelley, M. E., Han, N., Harris, A. W. F., Spencer, K. M., & Whitford, T. J. (2019). Inner speech is accompanied by a temporally-precise and content-specific corollary discharge. *NeuroImage*, *198*, 170–180. https://doi.org/10.1016/j.neuroimage.2019.04.038

Jirout, J., & Klahr, D. (2020). Questions – and some answers – about young children's questions. *Journal of Cognition and Development*, *21*(5), 729–753. https://doi.org/10.1080/15248372.2020.1832492

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Harvard University Press.

Johnson-Laird, P. N., Bucciarelli, M., Mackiewicz, R., & Khemlani, S. S. (2022). Recursion in programs, thought, and language. *Psychonomic Bulletin & Review*, *29*(2), 430–454. https://doi.org/10.3758/s13423-021-01977-y

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science*, *10*(3), 307–321. https://doi.org/10.1111/j.1467-7687.2007.00585.x

Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A Hierarchy of Time-Scales and the Brain. *PLOS Computational Biology*, *4*(11), e1000209. https://doi.org/10.1371/journal.pcbi.1000209

Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience*, *28*(47), 12224–12230. https://doi.org/10.1523/JNEUROSCI.3579-08.2008

Kikutani, M., Roberson, D., & Hanley, J. R. (2008). What's in the name? Categorical perception for unfamiliar faces can occur through labeling. *Psychonomic Bulletin & Review*, *15*(4), 787–794. https://doi.org/10.3758/PBR.15.4.787

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, *105*(31), 10681–10686. https://doi.org/10.1073/pnas.0707835105

Kittani, S. R., & Brinthaupt, T. M. (2024). Exploring self-talk in response to disruptive and emotional events. *Journal of Constructivist Psychology*, *37*(2), 129–143. https://doi.org/10.1080/10720537.2023.2194691

Kochanska, G., Murray, K., Jacques, T. Y., Koenig, A. L., & Vandegeest, K. A. (1996). Inhibitory control in young children and its role in emerging internalization. *Child Development*, *67*(2), 490–507. https://doi.org/10.1111/j.1467-8624.1996.tb01747.x

Kreiner, H., & Eviatar, Z. (2024). The sound of thought: Form matters–The prosody of inner speech. *Physics of Life Reviews*, *51*, 231–242. https://doi.org/10.1016/j.plrev.2024.10.006

Kross, E., Bruehlman-Senecal, E., Park, J., Burson, A., Dougherty, A., Shablack, H., Bremner, R., Moser, J., & Ayduk, O. (2014). Self-talk as a regulatory mechanism: How you do it matters. *Journal of Personality and Social Psychology*, *106*(2), 304–324. https://doi.org/10.1037/a0035173

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from* (Vol. 6). New York: Basic Books.

Landau, B., & Leyton, M. (1999). Perception, object kind, and object naming. *Spatial Cognition and Computation*, *1*(1), 1–29. https://doi.org/10.1023/A:1010073227203

LaTourrette, A., Chan, D. M., & Waxman, S. R. (2023). A principled link between object naming and representation is available to infants by seven months of age. *Scientific Reports*, *13*(1), 14328. https://doi.org/10.1038/s41598-023-41538-y

Lidstone, J. S. M., Meins, E., & Fernyhough, C. (2010). The roles of private speech and inner speech in planning during middle childhood: Evidence from a dual task paradigm. *Journal of Experimental Child Psychology*, *107*(4), 438–451. https://doi.org/10.1016/j.jecp.2010.06.002

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and causal text constraints. *Discourse Processes*, *34*(3), 259–280. https://doi.org/10.1207/S15326950DP3403_2

Lucca, K., Horton, R., & Sommerville, J. A. (2019). Keep trying!: Parental language predicts infants' persistence. *Cognition*, *193*, 104025. https://doi.org/10.1016/j.cognition.2019.104025

Lupyan, G. (2006). Labels facilitate learning of novel categories. In *The Evolution of Language* (pp. 190–197). WORLD SCIENTIFIC. https://doi.org/10.1142/9789812774262_0025

Lupyan, G. (2007). *Reuniting categories, language, and perception*. *29*(29).

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label-feedback hypothesis. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G. (2016). The centrality of language in human cognition. *Language Learning*, *66*(3), 516–553. https://doi.org/10.1111/lang.12155

Lupyan, G., Rahman, R. A., Boroditsky, L., & Clark, A. (2020). Effects of language on visual perception. *Trends in Cognitive Sciences*, *24*(11), 930–944. https://doi.org/10.1016/j.tics.2020.08.005

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

Lupyan, G., & Swingley, D. (2012). Self-directed speech affects visual search performance. *Quarterly Journal of Experimental Psychology*, *65*(6), 1068–1085. https://doi.org/10.1080/17470218.2011.647039

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: Activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology. General*, *141*(1), 170–186. https://doi.org/10.1037/a0024904

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences*, *110*(35), 14196–14201. https://doi.org/10.1073/pnas.1303312110

Luria, A. R. (1965). L.S. Vygotsky and the problem of localization of functions. *Neuropsychologia*, *3*(4), 387–392. https://doi.org/10.1016/0028-3932(65)90012-6

Lurito, J. T., Kareken, D. A., Lowe, M. J., Chen, S. H. A., & Mathews, V. P. (2000). Comparison of rhyming and word generation with fMRI. *Human Brain Mapping*, *10*(3), 99–106. https://doi.org/10.1002/1097-0193(200007)10:3%253C99::AID-HBM10%253E3.0.CO;2-Q

Mandler, J. M. (1984). *Stories, scripts, and scenes: Aspects of schema theory*. Psychology Press. https://doi.org/10.4324/9781315802459

Mar, R. A., & Oatley, K. (2008). The function of fiction is the abstraction and simulation of social experience. *Perspectives on Psychological Science*, *3*(3), 173–192. https://doi.org/10.1111/j.1745-6924.2008.00073.x

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.

McCarthy-Jones, S., & Fernyhough, C. (2011). The varieties of inner speech: Links between quality of inner speech and psychopathological variables in a sample of young adults.

Consciousness and Cognition, *20*(4), 1586–1593. https://doi.org/10.1016/j.concog.2011.08.005

McClelland, J. L., & Rumelhart, D. E. (1999). Distributed memory and the representation of general and specific information. In *Connectionist Psychology*. Psychology Press.

Mirza, M. B., Adams, R. A., Friston, K., & Parr, T. (2019). Introducing a Bayesian model of selective attention based on active inference. *Scientific Reports*, *9*(1), 13915.

Morin, A. (2005). Possible links between self-awareness and inner speech: Theoretical background, underlying mechanisms, and empirical evidence. *Journal of Consciousness Studies*, *12*(4–5), 115–134.

Morin, A. (2018). The self-reflective functions of inner speech: Thirteen years later. In *Inner Speech*. Oxford University Press. https://doi.org/10.1093/oso/9780198796640.003.0012

Morin, A. (2022). Self-reported inner speech illuminates the frequency and content of self-as-subject and self-as-object experiences. *Psychology of Consciousness: Theory, Research, and Practice*, *9*(1), 93.

Morin, A., Duhnych, C., & Racy, F. (2018). Self-reported inner speech use in university students. *Applied Cognitive Psychology*, *32*(3), 376–382. https://doi.org/10.1002/acp.3404

Mulvihill, A., Matthews, N., Dux, P. E., & Carroll, A. (2023). Task difficulty and private speech in typically developing and at-risk preschool children. *Journal of Child Language*, *50*(2), 464–491. https://doi.org/10.1017/S0305000921000945

Nedergaard, J. S. K., & Lupyan, G. (2024). Not everybody has an inner voice: Behavioural consequences of anendophasia. *Psychological Science*, *35*(7), 780–797. https://doi.org/10.1177/09567976241243004

Nedergaard, J. S. K., Wallentin, M., & Lupyan, G. (2023). Verbal interference paradigms: A systematic review investigating the role of language in cognition. *Psychonomic Bulletin & Review*, *30*(2), 464–488. https://doi.org/10.3758/s13423-022-02144-7

Nersessian, N. (2008). *Model-based reasoning in scientific practice*. Brill. https://doi.org/10.1163/9789460911453_005

Noorman, S., Neville, D. A., & Simanova, I. (2018). Words affect visual perception by activating object shape representations. *Scientific Reports*, *8*(1). Scopus. https://doi.org/10.1038/s41598-018-32483-2

Noss, R., & Hoyles, C. (1995). The dark side of the moon. In R. Sutherland & J. Mason (Eds), *Exploiting Mental Imagery with Computers in Mathematics Education* (pp. 190–201). Springer. https://doi.org/10.1007/978-3-642-57771-0_13

Oliveri, M., Finocchiaro, C., Shapiro, K., Gangitano, M., Caramazza, A., & Pascual-Leone, A. (2004). All talk and no action: A transcranial magnetic stimulation study of motor cortex activation during action word production. *Journal of Cognitive Neuroscience*, *16*(3), 374–381.

Oppenheim, G. M., & Dell, G. S. (2010). Motor movement matters: The flexible abstractness of inner speech. *Memory & Cognition*, *38*(8), 1147–1160. https://doi.org/10.3758/MC.38.8.1147

Orvell, A., Vickers, B. D., Drake, B., Verduyn, P., Ayduk, O., Moser, J., Jonides, J., & Kross, E. (2021). Does distanced self-talk facilitate emotion regulation across a range of emotionally intense experiences? *Clinical Psychological Science*, *9*(1), 68–78. https://doi.org/10.1177/2167702620951539

Parr, T., Corcoran, A. W., Friston, K. J., & Hohwy, J. (2019). Perceptual awareness and active inference. *Neuroscience of Consciousness*, *2019*(1), niz012.

Parr, T., & Friston, K. J. (2019). Generalised free energy and active inference. *Biological Cybernetics*, *113*(5), 495–513. https://doi.org/10.1007/s00422-019-00805-w

Paulesu, E., Frith, C. D., & Frackowiak, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, *362*(6418), 342–345. https://doi.org/10.1038/362342a0

Pecman, M. (2014). Variation as a cognitive device: How scientists construct knowledge through term formation. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, *20*(1), 1–24. https://doi.org/10.1075/term.20.1.01pec

Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: Evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain and Language*, *135*, 66–72. https://doi.org/10.1016/j.bandl.2014.05.005

Pezzulo, G., D'Amato, L., Mannella, F., Priorelli, M., Van de Maele, T., Stoianov, I. P., & Friston, K. (2024). Neural representation in active inference: Using generative models to interact with—and understand—the lived world. *Annals of the New York Academy of Sciences*, *1534*(1), 45–68. https://doi.org/10.1111/nyas.15118

Phillips, L. H. (1999). The role of memory in the Tower of London task. *Memory*, *7*(2), 209–231. https://doi.org/10.1080/741944066

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Pouw, W., Harrison, S. J., & Dixon, J. A. (2020). Gesture–speech physics: The biomechanical basis for the emergence of gesture–speech synchrony. *Journal of Experimental Psychology: General*, *149*(2), 391–404. https://doi.org/10.1037/xge0000646

Pratts, J., Pobric, G., & Yao, B. (2023). Bridging phenomenology and neural mechanisms of inner speech: ALE meta-analysis on egocentricity and spontaneity in a dual-mechanistic framework. *NeuroImage*, *282*, 120399. https://doi.org/10.1016/j.neuroimage.2023.120399

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*(7), Article 7. https://doi.org/10.1038/nrn1706

Pulvermüller, F. (2013). How neurons make meaning: Brain mechanisms for embodied and abstract-symbolic semantics. *Trends in Cognitive Sciences*, *17*(9), 458–470. https://doi.org/10.1016/j.tics.2013.06.004

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews Neuroscience*, *11*(5), 351–360. https://doi.org/10.1038/nrn2811

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793–797. https://doi.org/10.1111/j.1460-9568.2005.03900.x

Rabovsky, M., Sommer, W., & Abdel Rahman, R. (2012). Depth of conceptual knowledge modulates visual processes during word reading. *Journal of Cognitive Neuroscience*, *24*(4), 990–1005. https://doi.org/10.1162/jocn_a_00117

Reuter, T., Borovsky, A., & Lew-Williams, C. (2019). Predict and redirect: Prediction errors support children's word learning. *Developmental Psychology*, *55*(8), 1656–1665. https://doi.org/10.1037/dev0000754

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*(4), 378–411. https://doi.org/10.1016/j.cogpsych.2004.10.001

Ronfard, S., Zambrana, I. M., Hermansen, T. K., & Kelemen, D. (2018). Question-asking in childhood: A review of the literature and a framework for understanding its development. *Developmental Review*, *49*, 101–120. https://doi.org/10.1016/j.dr.2018.05.002

Roy, D. (2005). Semiotic schemas: A framework for grounding language in action and perception. *Artificial Intelligence*, *167*(1), 170–205. https://doi.org/10.1016/j.artint.2005.04.007

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Psychology Press. https://doi.org/10.4324/9780203781036

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *eLife*, *8*, e41703. https://doi.org/10.7554/eLife.41703

Scott, M. (2013). Corollary discharge provides the sensory content of inner speech. *Psychological Science*, *24*(9), 1824–1830. https://doi.org/10.1177/0956797613478614

Senay, I., Albarracín, D., & Noguchi, K. (2010). Motivating goal-directed behavior through introspective self-talk: The role of the interrogative form of simple future tense. *Psychological Science*, *21*(4), 499–504. https://doi.org/10.1177/0956797610364751

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565–573. https://doi.org/10.1016/j.tics.2013.09.007

Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Murray, R. M., & McGUIRE, P. K. (2001). A functional study of auditory verbal imagery. *Psychological Medicine*, *31*(2), 241–253. https://doi.org/10.1017/S003329170100335X

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology*, *23*(1), 117–141. https://doi.org/10.1348/026151004X21332

Smith, E. E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Sciences*, *95*(20), 12061–12068. https://doi.org/10.1073/pnas.95.20.12061

Sohn, M.-H., Goode, A., Koedinger, K. R., Stenger, V. A., Fissell, K., Carter, C. S., & Anderson, J. R. (2004). Behavioral equivalence, but not neural equivalence—Neural evidence of alternative strategies in mathematical thinking. *Nature Neuroscience*, *7*(11), 1193–1194. https://doi.org/10.1038/nn1337

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, *43*(6), 482–489. https://doi.org/10.1037/h0055479

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156.

Stanzione, C., & Schick, B. (2014). Environmental language factors in theory of mind development: Evidence from children who are deaf/hard-of-hearing or who have specific language

impairment. *Topics in Language Disorders*, *34*(4), 296–312. https://doi.org/10.1097/TLD.0000000000000038

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, *1*, 166. https://doi.org/10.3389/fpsyg.2010.00166

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *Cortex*, *77*, 1–12. https://doi.org/10.1016/j.cortex.2016.01.002

Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Harvard University Press.

Toms, M., Morris, N., & Ward, D. (1993). Working memory and conditional reasoning. *The Quarterly Journal of Experimental Psychology Section A*, *46*(4), 679–699. https://doi.org/10.1080/14640749308401033

Tullett, A. M., & Inzlicht, M. (2010). The voice of self-control: Blocking the inner voice increases impulsive responding. *Acta Psychologica*, *135*(2), 252–256. https://doi.org/10.1016/j.actpsy.2010.07.008

Vallotton, C., & Ayoub, C. (2011). Use your words: The role of language in the development of toddlers' self-regulation. *Early Childhood Research Quarterly*, *26*(2), 169–181. https://doi.org/10.1016/j.ecresq.2010.09.002

Varley, R. (2002). Science without grammar: Scientific reasoning in severe agrammatic aphasia. In M. Siegal, P. Carruthers, & S. Stich (Eds), *The Cognitive Basis of Science* (pp. 99–116). Cambridge University Press. https://doi.org/10.1017/CBO9780511613517.006

Virtue, S., Parrish, T., & Jung-Beeman, M. (2008). Inferences during story comprehension: Cortical recruitment affected by predictability of events and working memory capacity. *Journal of Cognitive Neuroscience*, *20*(12), 2274–2284. https://doi.org/10.1162/jocn.2008.20160

Vygotsky, L. S. (1987). *Thinking and speech. The collected works of Lev Vygotsky* (Vol. 1). Plenum Press.

Waxman, S. (2013). Building a better bridge. In M. R. Banaji & S. A. Gelman (Eds), *Navigating the Social World: What Infants, Children, and Other Species Can Teach Us* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199890712.003.0053

Weller, P. D., Rabovsky, M., & Abdel Rahman, R. (2019). Semantic knowledge enhances conscious awareness of visual objects. *Journal of Cognitive Neuroscience*, *31*(8), 1216–1226. https://doi.org/10.1162/jocn_a_01404

Westermann, G., & Mareschal, D. (2014). From perceptual to language-mediated categorization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634), 20120391. https://doi.org/10.1098/rstb.2012.0391

Willems, R. M., Labruna, L., D'Esposito, M., Ivry, R., & Casasanto, D. (2011). A functional role for the motor system in language understanding: Evidence from theta-burst transcranial magnetic stimulation. *Psychological Science*, *22*(7), 849–854. https://doi.org/10.1177/0956797611412387

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Winsler, A. (2009). Still talking to ourselves after all these years: A review of current research on private speech. In A. Winsler, C. Fernyhough, & I. Montero (Eds), *Private Speech, Executive Functioning, and the Development of Verbal Self-Regulation* (pp. 3–41). Cambridge University Press. https://doi.org/10.1017/CBO9780511581533.003

Winsler, A. E., Fernyhough, C. E., & Montero, I. E. (2009). *Private speech, executive functioning, and the development of verbal self-regulation.* Cambridge University Press.

Winsler, A., León, J. R. D., Wallace, B. A., Carlton, M. P., & Willson-Quayle, A. (2003). Private speech in preschool children: Developmental stability and change, across-task consistency, and relations with classroom behaviour. *Journal of Child Language*, *30*(3), 583–608. https://doi.org/10.1017/S0305000903005671

Woods, A., Jones, N., Alderson-Day, B., Callard, F., & Fernyhough, C. (2015). Experiences of hearing voices: Analysis of a novel phenomenological survey. *The Lancet Psychiatry*, *2*(4), 323–331. https://doi.org/10.1016/S2215-0366(15)00006-1

Yao, B. (2025). It's about time: Rhythmic foundations of inner thought. *Physics of Life Reviews*, *52*, 194–196. https://doi.org/10.1016/j.plrev.2025.01.003

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience*, *23*(10), 3146–3152. https://doi.org/10.1162/jocn_a_00022

Yao, B., Belin, P., & Scheepers, C. (2012). Brain 'talks over' boring quotes: Top-down activation of voice-selective areas while listening to monotonous direct speech quotations. *NeuroImage*, *60*(3), 1832–1842. https://doi.org/10.1016/j.neuroimage.2012.01.111

Yao, B., & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition*, *121*(3), 447–453. https://doi.org/10.1016/j.cognition.2011.08.007

Yao, B., & Scheepers, C. (2015). Inner voice experiences during processing of direct and indirect speech. In L. Frazier & E. Gibson (Eds), *Explicit and Implicit Prosody in Sentence Processing* (Vol. 46, pp. 287–307). Springer International Publishing. https://doi.org/10.1007/978-3-319-12961-7_15

Yao, B., & Scheepers, C. (2018). Direct speech quotations promote low relative-clause attachment in silent reading of English. *Cognition*, *176*, 248–254. https://doi.org/10.1016/j.cognition.2018.03.017

Yao, B., Taylor, J. E., & Sereno, S. C. (2022). What can size tell us about abstract conceptual processing? *Journal of Memory and Language*, *127*, 104369. https://doi.org/10.1016/j.jml.2022.104369

Yao, B., Taylor, J. R., Banks, B., & Kotz, S. A. (2021). Reading direct speech quotes increases theta phase-locking: Evidence for cortical tracking of inner speech? *NeuroImage*, *239*, 118313. https://doi.org/10.1016/j.neuroimage.2021.118313

Zimmermann, K., & Brugger, P. (2013). Signed soliloquy: Visible private speech. *The Journal of Deaf Studies and Deaf Education*, *18*(2), 261–270. https://doi.org/10.1093/deafed/ens072

Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences*, *18*(5), 229–234. https://doi.org/10.1016/j.tics.2014.02.008

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. https://doi.org/10.1037/0096-3445.135.1.1