

Learning to See Meaning in Visual Words and Objects

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
Doctor rerum naturalium (Dr. rer. nat.)
im Fach Psychologie

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
M.Sc. Alexander Enge
geboren am 09.11.1996 in Räckelwitz

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr. Julia von Blumenthal

Dekan der Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin
Prof. Dr. Dr. Christian Ulrichs

Gutachter*innen:

Tag der mündlichen Prüfung:

Zusammenfassung

Wie schafft es unser Gehirn, die Dinge zu verstehen, die unsere Augen sehen? Wir nehmen unsere Umwelt nicht nur als einen Mix aus Formen und Farben war, sondern sehen unterschiedliche Objekte mit bestimmten Namen und Funktionen. Außerdem können wir Schriftzeichen, auf Papier gedruckt oder auf einem Bildschirm angezeigt, benutzen, um Informationen über Zeit und Raum hinweg zu übertragen, wie zum Beispiel diese Worte, die Sie gerade lesen. In dieser Dissertation präsentiere ich drei Studien, die sich mit verschiedenen Aspekten des Lernens visuell-semantischer Information beschäftigen, also damit, wie wir Bedeutung aus visuellem Input extrahieren, und wie diese Bedeutung wiederum die visuelle Wahrnehmung beeinflusst. Ich gebe einen Überblick über das visuelle und semantische System im menschlichen Gehirn bei Erwachsenen und stelle eine Meta-Analyse von Studien vor, welche mittels funktioneller Magnetresonanztomographie (fMRI) die Hirnaktivität bei semantischer Verarbeitung bei Kindern gemessen haben. Dabei zeigte sich, dass bei semantischer Verarbeitung bei Kindern und Erwachsenen größtenteils dieselben Hirnareale aktiv sind und sich somit das semantische Netzwerk im Gehirn vom Kindes- zum Erwachsenenalter nur wenig zu verändern scheint. Danach präsentiere ich zwei experimentelle Studien über das Erlernen neuer visuell-semantischer Information. In der ersten dieser Studien haben wir mittels fMRI untersucht, wie sich die Hirnaktivität bei der Verarbeitung geschriebener und gesprochener Worte verändert, während Kinder lesen lernen. Dabei fanden wir Evidenz für nicht-lineare Veränderungen in der Hirnaktivität in verschiedenen Arealen des Neokortex sowie Hinweise auf eine Angleichung von visueller und auditiver Aktivität im ventralen okzipital-temporalen Kortex. In der letzten Studie haben wir untersucht, ob Erwachsene schnell die Funktion von zuvor unbekannten visuellen Objekten erlernen können und ob dieses semantische Wissen beeinflusst, wie die Objekte visuell wahrgenommen werden. Mittels Elektroenzephalographie (EEG) fanden wir heraus, dass das Verständnis für die Funktion eines Objekts nicht nur spätere, kognitive Prozesse (N400-Komponente) beeinflusst, sondern auch frühe Prozesse innerhalb der ersten 200 ms nach Präsentation des Objekts (P1- und N170-Komponente). Dies deutet darauf hin, dass semantische Information unmittelbar auf das visuelle System zurückwirken kann. Neben den Ergebnissen beschreibe ich die Methoden, welche für die hier beschriebenen Studien neu entwickelt wurden, in der Hoffnung, dass diese auch in zukünftigen Projekten nachgenutzt werden können. Abschließend diskutiere ich die Bedeutung der vorliegenden Ergebnisse sowie die Limitationen meiner Studien und der Forschung in diesem Bereich der kognitiven Neurowissenschaften im Allgemeinen.

Abstract

How does our brain make sense of what we see? We perceive the visual world around us not just as a mix of shapes and colors but as discrete objects that have names and functions. We can even use black lines and curves printed on paper or displayed on a computer screen to transmit meaningful information, as you are experiencing right now while reading these words. In this thesis, I present three studies that deal with different aspects of visual-semantic development, that is, how we process meaning based on visual input and how this semantic information feeds back to alter our visual perception. I review the visual and semantic system in adults and present a meta-analysis of functional magnetic resonance imaging (fMRI) studies of semantic cognition in children, showing that the semantic system seems to change relatively little between childhood and adulthood. I then present two experimental studies as examples of learning novel visual-semantic information. In the first of these, we longitudinally tracked brain activity (as measured using fMRI) in response to written and spoken words as children were learning to read, finding evidence for non-linear change in brain activity in different cortical areas and for an increase in audio-visual word processing in the left ventral occipito-temporal cortex. In the final study, we tested if adults can quickly learn the function of previously unfamiliar visual objects and if this semantic knowledge affects how these objects are perceived visually. We found that understanding the function of an object immediately affects visual brain responses (as measured using electroencephalography; EEG) within 200 ms after the presentation of the object (P1 and N170 components), as well as later higher-level processing (N400 component), indicating that semantic information rapidly feeds back into the visual system. I also provide details on some of the analysis methods developed for the present studies, with hope of making them reusable for future projects. Finally, I discuss the implications of my findings as well as limitations of the present studies and obstacles for scientific progress in this domain more broadly.

Contents

1	Introduction	1
1.1	Cognitive-neuroscientific methods	1
1.2	The visual system	3
1.3	The semantic system	4
1.4	Visual-semantic interfaces	6
2	Summary of the present studies	10
2.1	A meta-analysis of fMRI studies of semantic cognition in children	10
2.2	Tracking the neural correlates of learning to read with dense-sampling fMRI	12
2.3	Instant effects of semantic information on visual perception	14
3	General discussion	19
3.1	Visual-semantic processing in reading and object recognition	19
3.2	Implications for future research	22
3.3	Implications for real world applications	23
3.4	A need for better methods, questions, and theories	23
3.5	Conclusion	25
References		26
Original research articles		38
A meta-analysis of fMRI studies of semantic cognition in children		39
Tracking the neural correlates of learning to read with dense-sampling fMRI		63
Instant effects of semantic information on visual perception		93
Acknowledgements		123
Selbständigkeitserklärung		125

1 Introduction

In this thesis, I present three empirical studies on how humans learn to extract meaning from written words and visual objects. This question may seem like a very broad one and, as you will see, and as is so often the case in science, I cannot claim that I have found a clear and meaningful answer. Instead, I invite you to view these studies as three different examples of how visual-semantic development can be studied. Even though I cannot pretend that these studies were planned to build up on each other or logically follow from one another, I have always been interested in how our brain makes sense of the raw visual input that it gets, and how in turn this understanding might feed back into how we see the world around us. This has led me to pursue different research projects on different aspects of visual-semantic development throughout the course of my dissertation period. In this thesis, I present the outcomes of these different projects in the form of the original research articles as well as an introduction and discussion in which I provide an overview of the context of the different studies and how their results may inform one another as well as future research.

The thesis is structured as follows: In this Introduction, I will start out by introducing the cognitive-neuroscientific methods that are typically used to study visual and semantic processing non-invasively in humans. I will then briefly review common knowledge about the visual and semantic systems in the human brain as well as how they may interact with one another. I will then mention two specific examples of visual-semantic development, namely learning to understand written words and learning to understand visual objects. My review of the brain's semantic system forms the basis of the first research study included in this thesis, whereas the two examples of learning to understand written words and visual objects form the basis of the second and third research study, respectively.

Following the Introduction, I provide a brief summary of each of the three studies, including their theoretical background, research question, methods, and results. Finally, in the Discussion, I will summarize the main results of this thesis and how they interact with one another. I will also point out future research directions based on each of the different projects as well as their interactions. I will close with some more speculative remarks about what currently hampers us from gaining a deeper understanding about how the brain learns to make sense of the visual world.

Following the main text, the reader will find the reference section, a reprint of the three research studies, as well as an acknowledgment of the people who have supported me while working on this dissertation.

1.1 Cognitive-neuroscientific methods

Studying how the brain solves the task of making sense of visual information requires methods to measure brain activity while humans engage in tasks that require visual-semantic processing. This is not straightforward: For centuries, all that philosophers and scientists had access to was the input (What does a person see?) and the output (What does the person say or how do they behave?), whereas the actual processing inside the skull used to be a black box.¹ This changed in the last century, and especially in the last few decades, with the advent of non-invasive recording and imaging techniques, especially electroencephalography (EEG) and magnetic resonance imaging (MRI).

In EEG, electrodes are placed on the head of a human participant and connected to their scalp using an electroconductive gel. These electrodes are then able to pick up the electric fields created by the postsynaptic potentials of large neuronal populations along the cortical sheet of the brain. These measured amplitudes, which are typically on the order of a few microvolts, are a direct measure of neural activity (plus noise from the recording environment). Crucially, modern EEG devices can sample these amplitudes at a frequency of 500–2000 Hz (i.e., one sample every 2 to 0.5 ms), which is as high or higher than the timing of the postsynaptic potentials themselves, allowing researchers to track cortical information

¹An exception to this was the *post mortem* analysis of brain lesions and their association with certain behavioral particularities.

processing in real time. However, there are two major downsides of EEG recordings. The first is that they are very noisy, therefore making it necessary to collect and average a large number of trials to extract the signal of interest from background noise. The second is the low spatial specificity, since the voltage measured at any specific electrode is a complex mixture of signals that come from different areas of the brain (the “inverse problem”), due to volume conduction and different orientations of source dipoles depending on the folding of the neocortex. It is therefore difficult to impossible to link the EEG signal measured on the scalp to the specific neuronal populations having generated it. Nevertheless, EEG is an excellent tool to study the temporal sequence of cognitive processing in a comparatively low-cost and non-invasive fashion.

In MRI, the human participant is lying in an MRI scanner with a strong static magnetic field (typically on the order of 1.5–7 Tesla), which aligns the spins of Hydrogen protons in the brain. Then, a radio-frequency (RF) pulse is sent to push the protons into a higher level of magnetization. Once the RF pulse is turned off, the protons fall back into their original level (“relaxation”), emitting energy that gets picked up by a metal coil around the head. The timing of the relaxation depends on properties of the tissue, while specific additional magnetic gradients are used to encode the location of the signal being measured. This makes it possible to reconstruct an image of the local structural or functional properties of the brain based on the signal measured from each spatial unit (“voxel”) encoded via the magnetic field. In structural MRI (sMRI), the scanning sequence is created such that it measures differences in T1 or T2 relaxation times. This creates an image in which the intensity of the voxels depends on their fat content, which differs between gray matter (i.e., neuronal cell bodies), white matter (i.e., neuronal fibres), and cerebro-spinal fluid (CSF), thus providing a static high-resolution anatomical image of the brain. In functional MRI (fMRI), the scanning sequence measures the T2* relaxation time, which is sensitive to the blood oxygenation of the tissue. Since local neuronal activity requires oxygen, active brain areas receive an increase in blood flow and blood volume, which increases the local amount of oxyhemoglobin and decreases the amount of local deoxyhemoglobin, in turn leading to an increase in the measured T2* signal. By tracking the signal intensity at every voxel over time, that is, by reading out a whole-brain image of the T2* signal every couple of seconds, one can track the relative change in blood oxygenation over time, which is referred to as the blood oxygen level dependent (BOLD) effect. The BOLD effect is often interpreted as a measure of which brain areas are comparatively more or less active and how this activity covaries with a certain experimental manipulation (e.g., which brain areas show an increase in BOLD activity when processing faces as compared to houses). However, it is important to keep in mind that BOLD activity reflects changes in local blood oxygenation and is therefore only an indirect proxy of neuronal activity. Furthermore, the temporal resolution of fMRI is much worse than that of EEG, (a) because it typically takes approximately 2 s to sample every location in the brain and (b) because the haemodynamic response, that is, the increase in blood flow following an increase in local neuronal activity, takes a couple of seconds to unfold, even though the underlying neuronal activity happens on the order of microseconds. The spatial resolution, on the other hand, is much better than in EEG because fMRI measures each individual voxel inside the brain and therefore does not suffer from the inverse problem. Still, a typical voxel in fMRI is 8 to 27 mm³ in size and therefore contains at least multiple hundreds of thousands of neurons (Shapson-Coe et al., 2024), making it impossible to gain an insight into any neuronal circuit-level processing. Instead, fMRI can be used to identify which cognitive functions reliably engage which larger brain areas (think of the Brodmann areas or more modern whole-brain atlases, e.g., Destrieux et al., 2010; Glasser et al., 2016; Schaefer et al., 2018; Yeo et al., 2011), relative to some control condition.²

²Note, however, the common pitfall that when brain area *A* shows a significant change in BOLD activity while area *B* does not, this cannot be interpreted as area *A* being more active than area *B* without explicitly testing the *task × area* interaction effect (Nieuwenhuis et al., 2011).

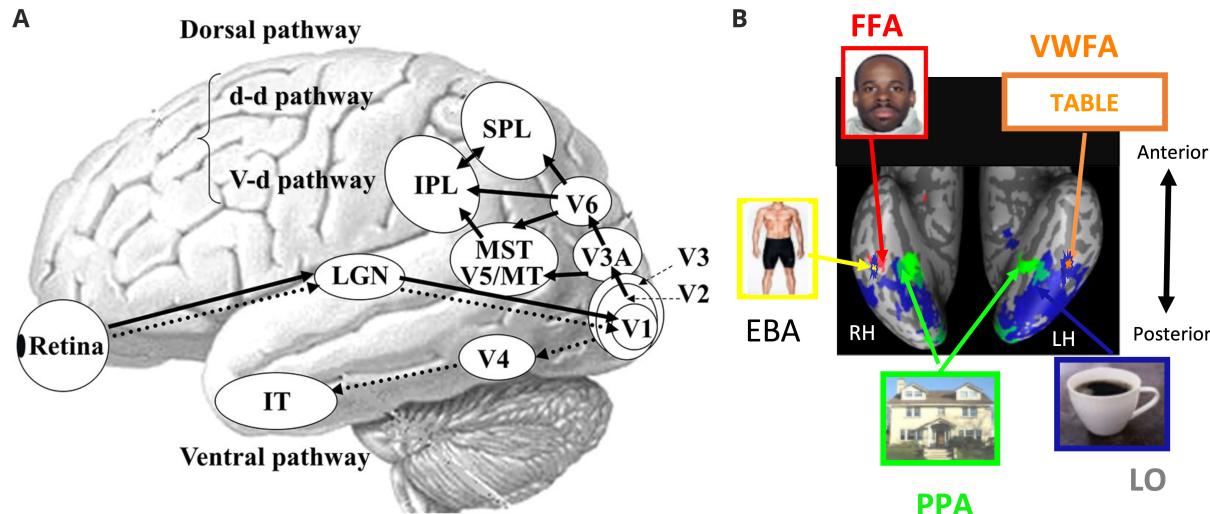


Figure 1. Visual processing areas in the human brain. A Schematic of the most important anatomical areas involved in visual processing in humans. The visual input from the retina is passed to the lateral geniculate nucleus (LGN) in the thalamus and then to areas V1 and V2 in the occipital lobe. From there, the dorsal (“where/how”) and ventral (“what”) visual streams represent the visual input in an increasingly abstract fashion. IPL = inferior parietal lobule, IT = inferior temporal cortex, MST = medial superior temporal area, MT = middle temporal area, SPL = superior parietal lobule. Adapted from Yamasaki and Tobimatsu (2018) under the terms of the Creative Commons Attribution (CC BY) License. B Object category-selective regions in the human brain. In the ventral and lateral occipito-temporal cortex, patches of fMRI voxels respond preferentially to certain object categories, including the fusiform face area for faces, the visual word form area (VWFA) for written words, the lateral occipital area (LO) for visual objects, the parahippocampal place area (PPA) for scenes, and the extrastriate body area (EBA) for bodies and body parts. Adapted from Behrmann and Plaut (2020) under the terms of the CC BY License.

1.2 The visual system

Vision is widely believed to be the most important sensory input modality in humans (Enoch et al., 2019; Winter et al., 2018) and occupies approximately one quarter of the human brain’s cortical surface, more than all other sensory modalities combined (Van Essen, 2003). The visual system processes the neural signals coming from the retinae of both eyes in a hierarchical sequence of subcortical and cortical areas, starting with the Lateral Geniculate Nucleus (LGN) in the Thalamus and continuing in the visual cortex in the occipital lobe of the brain (see Figure 1A; Felleman & Van Essen, 1991). There, Visual Areas 1 and 2 (V1 and V2) represent the visual field in a retinotopic fashion, with the firing rate of nearby cortical columns coding for the visual input from nearby locations in visual space. From V1 and V2, the input is passed on to two distinct but interacting processing streams, the dorsal stream, including areas V3, V5/IT, and parts of the parietal cortex, and the ventral stream, including areas V4 and parts of the ventral and lateral occipito-temporal cortex (e.g., the visual word form area [VWFA], the parahippocampal place area [PPA], and the fusiform face area [FFA]; see Figure 1B; Cohen et al., 2000; Epstein et al., 1999; Kanwisher et al., 1997). The dorsal stream predominantly codes for object location and object-related action guidance, whereas the ventral stream predominantly codes for object identity and categorization (Farivar, 2009; Mishkin & Ungerleider, 1982).

Within the ventral stream, subsequent areas have increasingly larger receptive field sizes, that is, they respond to increasingly large areas of real-world visual space. They also contain increasingly abstract representations of visual stimuli, disregarding low-level visual features such as size, luminance, or orientation, and thereby allowing the brain to classify a visual stimulus as a certain type of object and pass this information on to other neural systems for further processing and appropriate action selection. This process of “core object recognition” (DiCarlo et al., 2012) is often described as taking place fast (within less than 200 ms) and in a largely feed-forward fashion, progressing from LGN via early visual areas (V1 and V2) to higher-level visual areas (e.g., VWFA, PPA, FFA; DiCarlo et al., 2012; Felleman & Van Essen, 1991; Marr, 1982; Zhaoping, 2014). However, the visual system is highly interconnected and contains not only feed-forward axonal connections from earlier to later visual areas, but also horizontal

connections within visual areas and feedback connections from later to earlier visual areas (Lamme et al., 1998; Lamme & Roelfsema, 2000). These feedback connections help to improve visual perception especially under challenging conditions (e.g., occlusion and ambiguity; Hupé et al., 1998; Williams et al., 2008; Wyatte et al., 2014). Furthermore, recent evidence from human behavioral, EEG, and fMRI studies suggests that even early visual processing is modulated by non-visual information such as semantic knowledge, linguistic structures, and emotions (e.g., Abdel Rahman & Sommer, 2008; Boutonnet & Luyyan, 2015; Clarke et al., 2016; Hsieh et al., 2010; Phelps et al., 2016; Slivac et al., 2021; Teufel et al., 2018). Thus, while the understanding of bottom-up hierarchical visual processing has been one of the huge success stories of neuroscience (e.g., Hubel & Wiesel, 1962), and has successfully inspired advances in computer vision and artificial intelligence (e.g., He et al., 2015; Krizhevsky et al., 2012), it is clear that the visual system cannot be viewed in isolation: Its representations feed into other cognitive systems, such as the semantic or motor system, and are in turn influenced by them (Churchland et al., 1994).

1.3 The semantic system

Semantic processing refers to the storage, modification, and retrieval of knowledge about the world, including the meaning of spoken and written words, the significance and functions of objects, and abstract ideas or facts (e.g., that Paris is the capital of France). Humans can reason routinely about the taxonomic belonging of concepts to categories (e.g., a mouse is a mammal) and about the similarities and dissimilarities between different concepts (e.g., badminton is similar to tennis but without the ball touching the ground), to a degree that other animal species or machines seem unable to.³ However, it is important to note that semantic processing and related termini (e.g., “meaning,” “concept”) are often used somewhat vaguely and with different connotations in different scientific disciplines (e.g., philosophy, psychology, and linguistics; see Reilly et al., 2024). Here I will focus on representations of semantic knowledge and their neural correlates in humans as studied in psychology and cognitive neuroscience.

Broadly speaking, there are three different types of theories about the organization of semantic information in the human mind and brain (see Figure 2; Frisby et al., 2023):

- *Categorical* theories assume that exemplars (e.g., the written word “bird” or the picture or sound of a bird) are assigned to one of a set of discrete basic-level object categories or concepts (in this case, *bird*). The categories (e.g., *bird* versus *tree*) are typically mutually exclusive and defined by certain criteria (e.g., *has wings*) or via similarity to prototypical examples. On this view, concepts may be represented as vectors in a high-dimensional space, where each dimension is one category. In this space, each concept is a binary vector with a value of one on the category to which it belongs and zeros elsewhere. Concepts are then considered similar or dissimilar if they belong to the same versus to different categories.
- *Feature-based* theories assume that exemplars are recognized via the presence or absence of certain features. For instance, the concept of a bird is defined by the presence of features such as *lives*, *has wings*, *lays eggs*, etc., and these features get activated whenever a specific instance of the concept (e.g., the written word “bird” or the picture or sound of a bird) is encountered. Features can be binary (e.g., an animal *has wings* or not) or continuous (e.g., an animal *runs slow* versus *fast* versus *very fast* etc.), but they must be interpretable. On this view, the high-dimensional semantic space is formed by features instead of categories, and each concept is a vector that codes the presence or absence (for binary features) or the degree to which it possesses a certain feature (for continuous

³Regarding the comparison to non-human animals, some claim that the uniquely human advantage is having *language* rather than *semantics* (e.g., Friederici, 2017) but the concepts seem difficult to disentangle given that they are so tightly interlinked and correlated across phylogeny and ontogeny. For machines, semantic processing that would satisfy or impress us as humans seemed impossible just a few years before I started to write this thesis. Now, however, large language models based on the transformer architecture and huge training datasets can emulate or even outperform human semantic knowledge, although the degree to which they are able to “reason” or “understand” remains a source of fierce debate (e.g., Lewis & Mitchell, 2024; Sun et al., 2024; Wu et al., 2024).

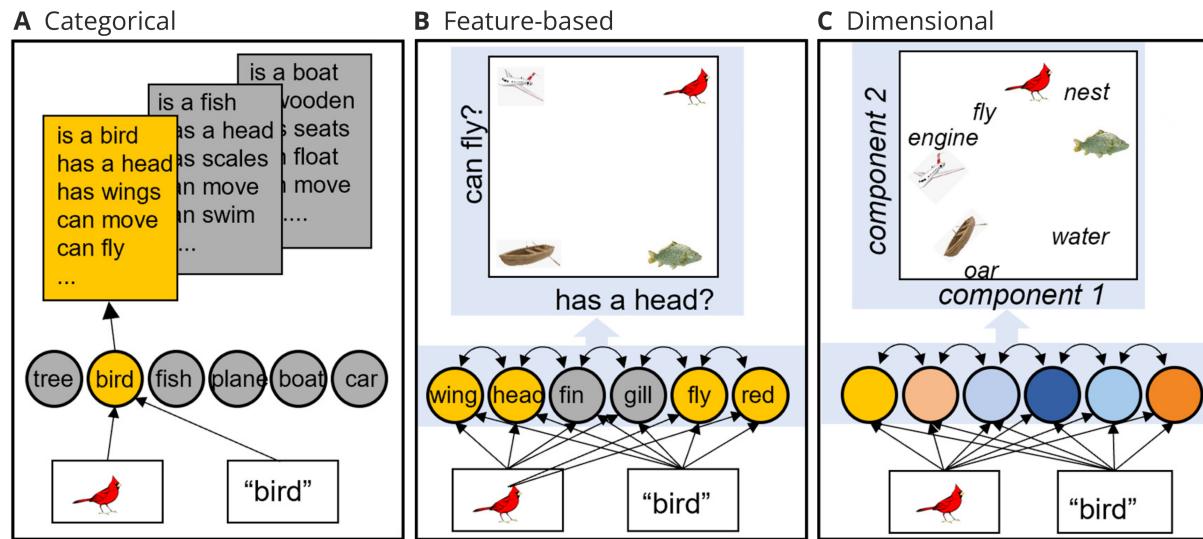


Figure 2. Theories of semantic representations. A Categorical theories assume that the sensory input is assigned to one of a set of discrete semantic objects categories or concepts. B Feature-based theories assume that concepts are defined by the presence or absence of certain semantic features. C Dimensional theories assume that concepts exists as points in an abstract, high-dimensional semantic space, the geometry of which needs not be humanly interpretable but still represents the semantic similarity and dissimilarity of objects. Adapted from Frisby et al. (2023) under the terms of the CC BY License.

features). Concepts are then considered similar or dissimilar if they show large versus little overlap in the presence/absence or degree of semantic features.

- *Dimensional* theories are like feature-based theories in that they assume concepts to be points in a high-dimensional semantic space and that the concept vectors may contain multiple non-zero values, coding for the degree to which a concept fits to a certain dimension. Unlike for feature-based theories, however, the dimensions need not be interpretable by humans in the same way a certain binary or continuous semantic feature like *has wings* is. Instead, the dimensions are typically learned by some computational model from a large set of input data, such as language models trained on large corpora of text from the internet or computer vision models trained on a large set of natural images. Even though the dimensions are not interpretable, the model captures the similarity and dissimilarity of concepts based on how near versus far away their distance in the high-dimensional semantic space is.

For each type, there are many different instantiations of specific theories (e.g., classical versus prototype-based categorical theories, dimensional theories based on word embeddings versus neural networks), and multiple of these theories may be implemented in the brain and used depending on the specific goal, task context, sensory input and output modality, and the availability of cognitive resources. It is important to note that the different theories often result in different experimental designs and analysis choices that are able to discover different behavioral and neural effects of semantic processing (Frisby et al., 2023). For instance, a categorical model is easily tested by running a highly controlled lab experiment that requires participants to make category judgments on visually presented words or images, and then look for differences in behavior or brain activity between different categories or objects (while ideally controlling as best as possible for differences in their low-level sensory features, e.g., Alizadeh et al., 2017; Rice et al., 2014). Dimensional models, on the other hand, are best tested by presenting a large volume of naturalistic input, such as natural speech or movies, and then using encoding models or decoding models to test which brain responses covary with the embedding vectors of a computational model trained on the same input (e.g., Huth et al., 2012, 2016).

Traditionally, the neural correlates of semantic processing in humans have been studied using univariate fMRI techniques, that is, by examining the difference in BOLD activity amplitude between two or

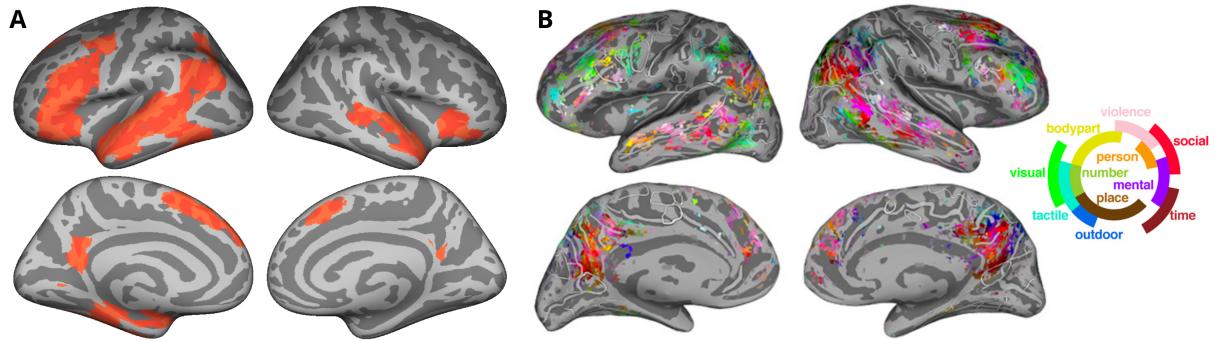


Figure 3. Semantic processing in the human brain. A Brain areas reliably activated by semantic task contrasts, based on an activation likelihood estimation (ALE) meta-analysis of 415 fMRI studies (for details, see Enge et al., 2021, and Section 2.1). Original data provided by Jackson (2021). B Brain areas representing specific semantic dimensions in a naturalistic task (story listening), estimated by fitting an encoding model of high-dimensional semantic representations of the words in the story to the time series of each voxel (for details, see Huth et al., 2016). Adapted from Deniz et al. (2019) under the terms of the CC BY License.

more conditions that are similar in their lower-level (e.g., sensory, phonological, orthographic) processing demands but differ in their semantic processing demands (e.g., reading words versus pseudowords). In addition to equating low-level processing, it also important to equate task difficulty and working memory demands between the different semantic conditions (Binder et al., 2009). Many individual studies and meta-analyses (e.g., Binder et al., 2009; Jackson, 2021; Rodd et al., 2015) have shown that these “semantic contrasts” reliably engage a relatively circumscribed, left-lateralized set of cortical regions (see Figure 3A), consisting of the inferior parietal lobe (especially the angular gyrus), lateral temporal cortex (especially the middle and inferior temporal gyrus), the ventral temporal cortex (fusiform and parahippocampal gyri), the dorso-medial prefrontal cortex, and the inferior frontal gyrus (especially the *pars orbitalis*). Additionally, the middle fusiform cortex and anterior temporal lobe (Forseth et al., 2018; Jackson et al., 2015; Visser et al., 2010; Woolnough et al., 2020) are reliably activated by semantic contrasts but are difficult to measure with fMRI due to the low signal-to-noise ratio in this area (Embleton et al., 2010; Liu, 2016). In the first empirical research article presented in this thesis (see Section 2.1), we investigated the development of this “localized” semantic system, that is, which brain areas are reliably activated by different types of semantic fMRI tasks in children and if these are similar or dissimilar to those in adults.

More recently, studies have emerged that moved beyond simple semantic task contrasts and tried to track semantic processing under more naturalistic circumstances, e.g., by scanning participants while viewing movies or listening to audio books or radio programmes (e.g., Deniz et al., 2019; Huth et al., 2012, 2016). One can then label the objects or words at each time point in the movie or audio stream, represent these concepts as vectors in a high-dimensional semantic space (e.g., from a word embedding or neural network model), and correlate the timeseries of these dimensions with the time series of the measured brain activity. Variants of these techniques usually show a somewhat different picture from the univariate semantic contrasts described above, with different concepts and semantic dimensions being scattered across almost the entire cortex, including areas that have typically not been thought to be engaged in semantic processing (see Figure 3B). While these different results clearly stem from the vastly different experimental designs and analysis choices (for a systematic discussion, see Frisby et al., 2023), there is not yet a clear path for integrating these different lines of research into an overarching theory of semantic processing in the human brain.

1.4 Visual-semantic interfaces

The visual and semantic systems need to interact with one another in order to make sense of what we see and to act accordingly. These interactions can be bidirectional: On the one hand, the visual system will feed its output (a high-level, feature invariant representation of the retinal input) into the semantic



Figure 4. Seeing without meaning. A How do these words sound and what do they mean? B What can be seen in this image? C Do you know what kind of object this is or what one can do with it? See Figure 5 for solutions as well as image rights and permissions.

system, allowing us, among many other options, to name visual objects, assess their usefulness to us, or relate them to other aspects of our cognition (e.g., our memory or current emotional state). On the other hand, the semantic system may modulate different stages of visual processing in a top-down fashion. This can be useful to predict visual input based on prior experience (as stipulated in predictive coding theories of the brain, e.g., A. Clark, 2013; Friston & Kiebel, 2009; Rao & Ballard, 1999) and to improve accuracy under the oftentimes challenging conditions in real-world vision (Hupé et al., 1998; Williams et al., 2008; Wyatte et al., 2014).

These bidirectional links between vision and semantics likely exist from very early stages of brain development onwards, allowing young infants to develop their first pieces of world knowledge based on their visual input. However, the ability to associate previously unknown visual shapes and objects with semantic information remains a crucial cognitive skill throughout our lifetime—or otherwise we would be lost whenever we try to understand any sign or gadget that we have never seen before. In the following, I will briefly introduce two examples of learning novel visual-semantic associations, namely children’s learning to read and adults’ learning to understand previously unknown visual objects. These two examples are examined more closely in the second and third empirical research articles presented in this thesis, respectively (see Sections 2.2 and 2.3).

1.4.1 Learning to see meaning in written letters. For most readers of this thesis, the black curves and lines displayed in Figure 4A will not bear much meaning. Your visual system will process them just fine, representing the visual input from each location in the retina into one coherent percept of black squiggles on a gray background. You may even guess correctly that these are strings of letters, and therefore you may get some non-zero response in your visual word form area (VWFA), which is tuned to recognize written words (see Section 1.2 and Figure 1). However, when asked to read these letter strings aloud, decide if they are words or non-words, or decide if they are related to one another or not, you would very likely be puzzled, unless you happen to know Hindi. This is because most readers will have been trained to read in a different writing system, most likely the Latin alphabet, which they can use effortlessly to convert written letters into spoken language and vice versa, as you will see when looking at the English translation of the Hindi words, shown in Figure 5A.

When looking at the Hindi words in Figure 4A, you find yourself in a similar situation as most beginning readers for their “native” script. Most children are explicitly taught how to read by their teachers and parents starting at approximately 5–6 years of age, which is much later than when most children pick up how to comprehend and produce spoken language (for review, see Skeide & Friederici, 2016). Before that, children might have acquired some purely visual familiarity with letters of their own script, and might recognize some words that are especially important to them (e.g., their own name). But by and large, strings of written letters such as the ones presented in Figure 5A will not be understandable for them and, while surely engaging the visual system, will not elicit any reliable activation in language areas or in



Figure 5. Seeing with meaning. A English translation of the Hindi/Devanagari words displayed in Figure 4A. B Original image from which the binarized “Mooney” image (see Mooney, 1957) in Figure 4B was generated. The binary image was adapted from Reining and Wallis (2024) under the terms of the CC BY License. The original image is from the THINGS database (Hebart et al., 2019) under the terms of the CC BY License. C Functional description of the object displayed in Figure 4C, which was adapted with personal permission from Tim Daniels, <https://poultrykeeper.com>.

the semantic system. The process of learning to read can be thought of as fine-tuning the visual system towards the recognition of written letters (e.g., Dehaene-Lambertz et al., 2018), but even more so as connecting the visual system to the language and semantic systems, such that a previously meaningless letter string can be converted into speech sounds (e.g., during reading aloud) and word meaning (e.g., during lexical decision or text comprehension). Early during this process, beginning readers will rely more heavily on converting individual letters into speech sounds and merging these speech sounds into a spoken word form, the meaning of which they can then be accessed based on their already fully developed spoken language capacity (Ehri, 2005; Frith, 1986; Skeide & Friederici, 2016). After gaining more reading experience, children learn to skip the detour via this phonological (alphabetic) route and become able to read in an orthographic fashion, that is, process familiar words as a whole and access their meaning from the mental lexicon directly, without having to engage into individual grapheme–phoneme conversion.⁴ In the second study included in this thesis (see Sections 2.2), we tracked this process of visual-phonological and visual-semantic development by repeatedly measuring fMRI responses to spoken and written words in a sample of children that underwent approximately 1.5 years of reading instruction.

1.4.2 Learning to see meaning in visual objects. Our capacity to learn to connect previously meaningless visual information with semantic concepts luckily does not end during childhood. Instead, we remain able to form new links between the visual and semantic systems for our entire lives, allowing us to understand what kind of object or visual scene we are seeing (i.e., a link from vision to semantics) and to see the visual word around us differently based on what we expect or know to be true (i.e., a link from semantics to vision). As an example of this, take a look at Figure 4B. Can you tell what is displayed in this black and white image? If you cannot, take a look at Figure 5B, which shows the original image from which the black and white image was created (using binarization; see Mooney, 1957; Reining & Wallis, 2024). Once you return to the black and white image, you most likely cannot “unsee” the central object of the image anymore—your prior experience and knowledge of what is displayed has changed your perception, even though the visual input is the same as before (for formal studies of this phenomenon using the same kind of stimuli, see Samaha et al., 2018; Teufel et al., 2018).

Another example of such feedback effects from semantic cognition to visual perception, and one that might bear more real-world applicability than the previous one, is the processing of visual objects and tools based on their function. We learn the names and functions of most visual objects as children but, in our ever more technology-driven world, continue to be frequently presented with tools and gadgets that

⁴Note that these two “routes” of reading, that is, an indirect connection from vision via phonology to semantics and a direct route from vision to semantics, are also part of computational cognitive models of reading (e.g., Coltheart et al., 2001; Perry et al., 2007, 2010, 2013; Seidenberg & McClelland, 1989).

we might have never seen before⁵. Learning what these objects are for or how to use them often makes them appear in a different light: We may get a kind of “aha” experience from understanding what we see and this may change how we perceive the objects and their visual features. For example, most people will not have encountered the object in Figure 4C at any point in their life, and will not have an intuition what this object is for or how to use it. However, after being informed about the function of the object (e.g., by taking a look at Figure 5C), some of the visual features of the object and their configuration start making sense. I would argue that, as in the previous example of the black and white image, we cannot “unsee” the meaning (i.e., function or usage) of this objects after we have discovered it. That is, learning the semantic meaning of a previously unfamiliar object might change our visual perception of the object, both while we are learning it (i.e., during the “aha” experience) and after we have learnt it (i.e., when we re-encounter the visual object later on). In the third study included in this thesis (see Section 2.3 and the original research article), we tested this idea empirically by presenting human adults with previously unfamiliar objects, as in the example above, and measuring their EEG response to the objects before, during, and after they learnt about the their functional meaning.

⁵Just think about fidget spinners, smart watches, or massage guns.

2 Summary of the present studies

2.1 A meta-analysis of fMRI studies of semantic cognition in children

In this first⁶ study (Enge et al., 2021), we provide a quantitative synthesis of the brain areas involved in semantic processing in children. To this end, we conducted a meta-analysis of all available fMRI studies in children that probed different aspects of semantic cognition, including semantic knowledge (e.g., naming an object after hearing a verbal description), semantic relations (e.g., deciding if a pair of words belong together or not), and visual object categorization (e.g., viewing object images from different semantic categories).

Meta-analysis is a useful tool to synthesize the research literature within a relatively narrow domain (Glass, 1976; Harrer et al., 2021; Hedges & Olkin, 1985). The surfacing of the latest replicability crisis in psychological science (e.g., Open Science Collaboration, 2015) was a reminder that the findings from any individual research paper cannot be accepted as true and irrevocable facts, due to problems such as small sample sizes, publication bias (the “file drawer effect,” i.e., the low publication rate of negative findings), *post hoc* theorizing, and questionable research practices (Ioannidis, 2005; Kerr, 1998; Simmons et al., 2011). To get a slightly more reliable and balanced view of the research outcomes in a certain area, a meta-analysis empirically integrates the findings from many—ideally all—studies on a given topic. By doing so, a meta-analysis can (a) show which effects reported in the original papers are likely to be robust and replicable, thereby reducing the number of false positives, and (b) identify novel effects that individual papers might not have had the statistical power to detect, thereby reducing the number of false negatives (Schmidt & Hunter, 2015).

In the fMRI literature, the evidential status of the results of individual papers is especially doubtful. This is because, compared to behavioral and neuroscientific studies using “cheaper” methods (e.g., EEG), data collection is at least an order of magnitude more costly and depends on the availability of specialized hardware and expertise. This typically leads to smaller sample sizes and, therefore, lower statistical power and less reliable estimates (Button et al., 2013; Szucs & Ioannidis, 2020; Szucs & Ioannidis, 2017). At the same time, data preprocessing and analysis methods are highly complex, which introduces many potential sources of error and researcher degrees of freedom. Within the fMRI literature, developmental studies are especially prone to these problems because it is more difficult to recruit children as compared to, e.g., psychology undergraduate students in need of course credit or monetary compensation. Furthermore, children tend to show lower levels of compliance and higher levels of head motion, which reduces the amount of data that can be collected as well as its quality.

Our goal was therefore to identify and meta-analyze all available fMRI studies on semantic cognition in children, building up on previous fMRI meta-analyses on semantic cognition in adults (Binder et al., 2009; Jackson, 2021; Rodd et al., 2015; Vigneau et al., 2006; Visser et al., 2010) as well as on our own previous fMRI meta-analysis on language processing in children (Enge et al., 2020). Using three online databases, we identified approximately 1,000 articles based on our keyword search, 45 of which fulfilled all of our inclusion criteria. Contained in these 45 articles were 50 fMRI experiments on semantic cognition in children, which we further classified into semantic knowledge experiments (21 experiments), semantic relatedness experiments (16 experiments), and visual object categorization experiments (13 experiments). In total, these experiments included data from 1,018 children with a mean age of 10.1 years (range = 4–15 years).

The experiments reported 687 peak voxel coordinates from statistically significant clusters associated with semantic cognition in children. Using the activation likelihood estimate (ALE) algorithm (Eickhoff et al., 2009; Eickhoff et al., 2012; Turkeltaub et al., 2002) implemented in the NiMARE software package (Salo et al., 2023), these activation peaks were convolved with a Gaussian smoothing function, integrated

⁶Note that order in which the studies are presented in this thesis was based on their content and logical flow, not on their chronological order of publication.

into a meta-analytic map, and thresholded using a permutation-based family-wise error correction for multiple comparisons at the whole-brain level (Eickhoff et al., 2012, 2016). For general semantic cognition, combined across all three task types, there was reliable BOLD activity across experiments in eight clusters, namely in the left inferior frontal gyrus (pars triangularis; two clusters), the bilateral supplementary motor area, the left fusiform gyrus, the right insula, the left middle temporal gyrus, the right inferior occipital gyrus, and the right fusiform gyrus. Broken down by task category, three clusters (left inferior frontal gyrus, bilateral supplementary motor area, and right insula) were activated for both semantic world knowledge experiments and semantic relatedness experiments, whereas the left middle temporal gyrus was only activated for semantic relatedness experiments and the left and right fusiform and occipital clusters were only activated for semantic object categorization experiments.⁷

We validated our results from the ALE analysis using a second meta-analytic approach (seed-based d mapping; Albajes-Eizagirre, Solanes, & Radua, 2019; Albajes-Eizagirre, Solanes, Vieta, et al., 2019), which gave qualitatively similar results and also allowed us to control for various experiment-level covariates (e.g., language, sensory modality, response modality).

We tested for age related effects both within our meta-analytic sample of children's experiments and by comparing our results in children with those from a previous meta-analysis on semantic cognition in adults (Jackson, 2021). Regarding age-related changes during childhood, we found one cluster in the right insula that showed significantly more reliable activation in older as compared to younger children (but note the limited statistical power of this analysis as well the statistical problems associated with median split analyses; e.g., Irwin & McClelland, 2003; McClelland et al., 2015). Regarding the comparison to adults, there was very large overlap across left and right inferior frontal, supplementary motor, and left temporal regions, but also significantly less reliable activation in children in the left anterior temporal lobe and significantly more reliable activation in fusiform and occipital regions, especially in the right hemisphere.⁸

Finally, we used two previously established empirical techniques to examine the robustness of our meta-analytic results against spurious effects and publication bias in the original literature (leave-one-out and fail-safe N ; Acar et al., 2018; Enge et al., 2020; Samartsidis et al., 2020). These analyses showed that all of our clusters from the main analysis and almost all clusters from the sub-analyses of task types were reliable, even if there were spurious original experiments or strong publication bias.

Taken together, our meta-analysis is the first of fMRI experiments on semantic cognition in children, and showed strong evidence for a reliably activated semantic network that is remarkably similar to those reported in the adult fMRI literature (for meta-analysis, see, e.g., Binder et al., 2009; Jackson, 2021). We found that experiments with linguistic materials (semantic world knowledge and relatedness judgments) reliably activated the same regions in the left inferior frontal and bilateral premotor cortices, whereas visual object categorization experiments activated a distinct set of regions in posterior brain areas (bilateral fusiform and right occipital cortices).⁹ Despite the large overlap between our meta-analytic results and previous meta-analytic results in adults (Jackson, 2021), there were a few statistically significant differences. The most striking of these was that there was significantly less reliable activation in children in the anterior part of the left temporal lobe, an area associated with very high-level and amodal semantic processing in adults (e.g., Lambon Ralph et al., 2017; Patterson et al., 2007). This difference may point to the fact that this "semantic hub" may not yet be fully developed in children at the age included in our meta-analysis (mostly 8–12 years; see Figure 3 in the original paper).

⁷However, note that this *difference in statistical significance* does not automatically imply a *statistically significant difference* (Gelman & Stern, 2006; Nieuwenhuis et al., 2011). For a formal statistical comparison of the different task types, see Figure 6 and Table 4 in the original paper.

⁸But note that this latter effect may be due to differences in the relative number of linguistic versus visual object categorization experiments included in the two meta-analyses.

⁹Potentially casting doubt on our decision to lump the latter type of experiments together with the former into one meta-analysis, as visual object category processing may be based purely on differences in lower-level visual features and may not necessarily engage higher-level semantic processing.

We have shared the data and analysis code of our meta-analysis on the Open Science Framework (<https://osf.io/34ry2>) and on GitHub (https://github.com/SkeideLab/meta_semantics) so that they can be scrutinized and reused.¹⁰

2.2 Tracking the neural correlates of learning to read with dense-sampling fMRI

In this second study (Enge & Skeide, 2024), we focused on children's learning to read as one case example for the human brain learning to associate visual shapes with semantic as well as phonological information. Learning to read is a major step in most people's cognitive development and is of great importance for the rest of their lives (for instance, allowing you to engage with this PhD thesis). From a phylogenetic point of view, reading is a fairly recent phenomenon. Therefore, it is not at all obvious that most human beings nowadays at some point in their lives become able to absorb and transmit semantic information via small and somewhat arbitrary black symbols printed on the page of a book or rendered on a digital screen. Yet, this is exactly what happened over the course of the last few millennia, rapidly picking up over the past ~500 years, and in turn fueling many other achievements of humankind, such as science and technology.

Unlike reading and writing, spoken language has been part of human culture and everyday life for hundreds of thousands of years, which left enough time to evolve dedicated brain hardware and neural mechanisms for spoken language comprehension and production (for review, see, e.g., Fedorenko et al., 2024; Ferstl et al., 2008; Friederici, 2017; Hagoort, 2016; Hickok & Poeppel, 2007). It is therefore unsurprising that human infants pick up spoken language comprehension quickly and relatively effortlessly within typically less than 2 years of age (for review, see Friederici, 2006; Skeide & Friederici, 2016), with certain aspects of speech processing already functioning *in utero* (e.g., Ghio et al., 2021). Reading, on the other hand, requires years of explicit instruction and is typically only learned during kindergarten and primary school age (typically 5–8 years). During the process of learning to read, a child's brain needs to become able to (a) reliably recognize letters in one's own writing system and distinguish them from other kinds of visual objects, (b) become able to link individual letters to their corresponding speech sounds to enable decoding of written words, and (c) as the child becomes more proficient, learn to access the meaning of an entire word from its visual word form (Ehri, 2005; Frith, 1986). While the first aspect (visual word form recognition) has received much attention in the cognitive neuroscience of reading and reading development (e.g., Cohen et al., 2000; Dehaene & Cohen, 2011; Dehaene-Lambertz et al., 2018), the two following processes of linking visual words to spoken language and linking visual words to word meaning have received much less attention (but see, e.g., Vin et al., 2024).

Studying the neural correlates of learning to read requires to take repeated measurements of brain activity (e.g., using EEG or fMRI) while children are participating in reading instruction. Doing so is often difficult for a number of reasons, including the fact that reading instruction is typically confounded with other aspects of schooling and general cognitive development, the difficulty of recruiting children and their families at the right age and for a prolonged period of time, the difficulty of collecting sufficiently large quantities of high-quality data (e.g., due to children's lower attention span and increased head motion during MRI scanning), and a lack of readily available statistical models and toolboxes for longitudinal analysis of neuroimaging data. Additionally, available longitudinal studies of the neural correlates of learning to read (for review, see Chyl et al., 2021) have exclusively been carried out with children from the Global North, which limits their generalizability to other cultural and socio-economic backgrounds as well as to different, non-alphabetic writing systems.

Our goal was to overcome these limitations and investigate the change in children's brain responses to written and spoken words as they are learning to read, and to do so in a cultural setting and writing system that has typically been neglected by developmental cognitive neuroscience. We expected that learning to read would cause an increase in brain activity in brain areas known to be involved in written

¹⁰At the time of writing, three new meta-analyses have acknowledged reusing our scripts (Bortolini et al., 2024; Cui et al., 2022; Yang et al., 2024).

and spoken word processing (e.g., the visual word form area). We also expected that the multi-voxel response patterns in audio-visual processing areas (e.g., the left posterior temporal sulcus; pSTS) would become more similar between written and spoken words as the brain learns to connect the former to the latter. Finally, we expected that responses to written words would become more similar to each other, that is, more stable from one scanning session to the next, as the brain becomes more finely tuned to visual letters.

We acquired fMRI from two groups of children in the region of Uttar Pradesh, India, which had no access to public schooling and received approximately 1.5 years of reading instruction (intervention group) or math instruction (active control group) as part of our study. For the purpose of the present paper, only the data from the reading instruction group was analyzed. The participants completed up to 6 experimental sessions, spaced at irregular intervals of approximately 2–3 months. During these sessions, we acquired a structural MRI scan, an fMRI scan, and behavioral data on standardized tests of reading skills, maths skills, working memory, and general cognitive ability. During the fMRI scan, children were presented with short blocks of written words, written pseudowords, and written low-level control stimuli (false fonts), as well as spoken words, spoken pseudowords, and spoken low-level control stimuli (noise-vocoded speech). We analyzed these longitudinal data using linear mixed-effects models to test for linear and non-linear changes in behavioral test scores, BOLD activity amplitude, multi-voxel audio-visual pattern similarity, and multi-voxel within-condition pattern stability. For this purpose, we created a novel implementation of whole-brain linear mixed-effects model using the high-performance programming language Julia (Bezanson et al., 2017), which we hope may be re-used for future longitudinal fMRI studies.¹¹

In the behavior data, we found a statistically significant improvement in test scores over the course of the study for most subtests of reading skills, including word and pseudoword reading accuracy, phoneme replacement, semantic fluency, reading comprehension, and dictation. Additionally, there also was a statistically significant improvement in two subtests of maths skills and in one subtest of working memory (backward digit span). The other subtests of reading skills (picture naming, verbal fluency, and word, pseudoword, and passage reading time) showed no significant change, as did the other subtests of working memory (Corsi block test forward and backward, backward digit span) and the test for general cognitive ability (Raven's Progressive Colored Matrices).

In the analysis of whole-brain BOLD activity amplitudes, we found robust sensory activation in the auditory cortex (bilateral superior temporal gyri and sulci) for all auditory conditions (spoken words, pseudowords, and low-level controls) and in the visual cortex (from early visual cortex to the ventral occipito-temporal cortex) for all visual conditions (written words, pseudowords, and low-level controls). However, there were only very few and small clusters that showed a statistically significant change (linear or non-linear) in BOLD activity amplitude over the course of the study. There were some non-linear changes for spoken stimuli in the auditory cortex and for written stimuli in the visual cortex, but their shape did not follow a pattern that we had predicted.¹² For the contrast of written words versus written low-level controls, there was one cluster at the ventral posterior cingulate cortex that showed an increase in BOLD activity over the first approximately 6 months of the study, followed by a decline over the remaining approximately 10 months, as we had predicted. However, this cluster was very small, the individual BOLD activity curves of individual participants in this cluster were very variable, and this anatomical location has typically not been considered as part of the visual word form recognition or language networks, therefore casting doubt on the reliability and meaningfulness of this finding.

In the analysis of multi-voxel audio-visual pattern similarity, we found that over the course of the study, BOLD activity patterns in the left ventral occipito-temporal (vOT) cortex became more similar over

¹¹For previous implementations of whole-brain linear mixed-effects models using the somewhat slower programming language R, see Chen et al. (2013) and Madhyastha et al. (2018).

¹²We had predicted a linear increase of BOLD activity or a non-linear, negative quadratic pattern (i.e., an inverted “u” shape), based on a previous study (Dehaene-Lambertz et al., 2018) and based on the expansion and renormalization model of brain plasticity (Wenger et al., 2017). What we observed in most clusters was a positive quadratic pattern (i.e., a “u” shape), which seems difficult to interpret in the context of brain plasticity and skill development.

time for the pair of contrasts of spoken words versus low-level controls and written words versus low-level controls. This was in line with our hypotheses and may indicate that over the course of learning to read, the brain becomes able to access phonological and semantic information from written words that it had previously only been able to access from spoken words. There was no significant change in audio-visual pattern similarity for any of the other pairs of contrasts and regions of interests.

In the analysis of multi-voxel within-condition pattern stability, we found that over the course of the study, BOLD activity patterns became less stable (i.e., self-similar) for auditory pseudowords and auditory words in the bilateral posterior superior temporal sulcus (pSTS). This was not in line with our *a priori* hypotheses, which predicted an *increase* in pattern stability specifically for written words and pseudowords.

Taken together, our study provides only weak evidence for the idea that learning to read increases BOLD activity amplitude and audio-visual processing in visual and language-related brain areas. While we did find some evidence for longitudinal change in BOLD activity amplitudes, these changes typically did not follow a theoretically predicted pattern of linear or non-linear growth and did not typically occur in areas associated with reading or audio-visual processing. However, we did find some evidence for an increase in audio-visual processing activity in the left vOT cortex. There was no reliable change in pattern stability, that is, we did not find that reading leads to more stable (i.e., self-similar) written word representations.

It is important to point out that these weak results may be driven by methodological shortcomings of our study, including a very small sample size (15 children with 2–6 sessions each; see Button et al., 2013; Ioannidis, 2005), an intervention that might not have been long enough to capture the entire process of becoming a sufficiently fluent reader (especially in the relatively complex Devanagari writing system), and an fMRI block design that was not suitable for multivariate analysis at the single item level, thereby preventing us from comparing individual word representations. We nevertheless hope that some of our findings can be replicated and that our general study design and longitudinal statistical analysis approach inspires future longitudinal work in different cultures and writing systems.

2.3 Instant effects of semantic information on visual perception

In this third and final study (Enge et al., 2023), we switched our focus from the development of visual-semantic processing during childhood to the learning of visual-semantic information in adults. Although most of our learning of the meaning of visual objects takes place during infancy (e.g., by learning to differentiate high-level visual object categories) and childhood (e.g., by learning to associate abstract visual shapes with meaning during learning to read), this capacity luckily remains available as we age. Otherwise, we would not be able to understand much at all in a world where we are presented with novel gadgets and symbols almost on a daily basis.

Our goal was to capture the moment during which we learn to understand the function of a previously unknown visual object, as well as to test if this understanding changes the way in which we perceive the object visually. This latter question has sparked endless debate among cognitive scientists. Some argue that perception and cognition are two distinct processing stages and that perception takes place in a feed-forward, modular fashion, and then passes its outputs on to other higher-level systems (e.g., those processing semantic information or emotions) but cannot itself be influenced by them (e.g., DiCarlo et al., 2012; Firestone & Scholl, 2016; Fodor, 1983, 1984; Machery, 2015; Pylyshyn, 1999). Others argue that this distinction between perception and cognition cannot be maintained and/or that cognition interacts with perception in a top-down fashion from the earliest stages of sensory processing on (e.g., Ahissar & Hochstein, 2004; Churchland et al., 1994; A. Clark, 2013; Friston & Kiebel, 2009; Lupyan et al., 2020; Thierry, 2016; von Helmholtz, 1867; Yuille & Kersten, 2006). Recent empirical evidence by and large supports the latter view. It has been shown that emotional, linguistic, and semantic information can change the response to visually identical stimuli in terms of behavioral measures (e.g., Gauthier et al., 2003; Phelps et al., 2016; Slivac et al., 2021), the BOLD response in visual cortex (e.g., Clarke et al.,

2016; Hsieh et al., 2010), and early ERP components in the EEG (e.g., Abdel Rahman & Sommer, 2008; Boutonnet & Lupyan, 2015; Samaha et al., 2018). The latter type of evidence seems especially informative because the high temporal resolution of EEG (on the order of microseconds, directly capturing cortical postsynaptic potentials) allows to test how early during processing one can detect high-level linguistic or semantic influences (for an extended version of this argument, see Athanasopoulos & Casaponsa, 2020). For instance, our research group and others have shown that letting participants learn semantic information about previously unknown visual stimuli elicits not only differences in relatively late, semantic ERP components (e.g., the N400 component), but also in early ERP components typically associated with visual perception (e.g., the P1 component; Abdel Rahman & Sommer, 2008; Boutonnet & Lupyan, 2015; Maier & Abdel Rahman, 2019; Samaha et al., 2018; Weller et al., 2019). However, all of these previous studies have used an extensive training phase in which participants learned to associate the visual stimuli with semantic information, whereas the EEG was only measured and analyzed after this learning had taken place.

In our study, we wanted to test if learning semantic information about previously unknown visual objects affects visual perception instantly, that is, directly as this information is being acquired and the function of the object is being understood. To this end, we presented participants with images of existing but rare visual objects (e.g., a galvanometer) in three separate phases. In the first phase, we presented the objects without any information, to probe if they were indeed unknown to participants. In the second phase, we presented a brief verbal description before the object that allowed participants to understand what kind of object they were seeing. In the third phase, we again presented the objects without any information, to test for downstream effects of the understanding acquired in the second phase. Crucially, in the second phase, we presented half of the objects with the correct, matching verbal descriptions, which typically allowed participants to form a correct understanding of the object, and the other half of the object with an incorrect, non-matching verbal description, which typically precluded such an understanding.¹³ This allowed us to compare responses to the same visual objects with and without a semantic understanding, and to do so before, while, and after this understanding had happened.

During all three phases of the experiment, we measured the EEG and analyzed three different ERP components, namely the P1 component (100–150 ms after stimulus onset) as a marker of early visual perception, the N170 component (150–200 ms) as a marker of high-level visual perception, and the N400 component (400–700 ms) as a marker of semantic processing. In the first phase, we observed no reliable differences in any of the three ERP components between objects that would subsequently show semantically informed perception as compared to semantically uninformed perception. This was expected given that the relevant semantic information had not yet been presented, so that all objects were unfamiliar and not understood by the participants. In the second phase, when the objects were presented with the semantic information (either a matching description, inducing semantically informed perception, or a non-matching description, keeping the perception semantically uninformed), we did observe reliable difference in two of the three ERP components. That is, semantically informed perception caused significantly larger (i.e., more negative) ERP amplitudes in the N170 component and significantly reduced (i.e., less negative) ERP amplitudes in the N400 component. There was no effect of semantically informed perception in the P1 component. In the third phase, when objects were presented once more without any information, semantically informed perception was associated with significantly larger (i.e., more positive) ERP amplitudes in the P1 component and, again, significantly reduced (i.e., less negative) ERP amplitudes in the N400 component.

We interpret these ERP effects as evidence that semantic information not only affects late, post-perceptual stages of visual object processing (i.e., the N400 component), but also earlier, perceptual processing (i.e., the P1 and N170 components). This conceptually replicates previous studies from our lab and others (e.g., Abdel Rahman & Sommer, 2008; Boutonnet & Lupyan, 2015; Maier & Abdel Rahman, 2019; Samaha et al., 2018; Weller et al., 2019) and provides evidence for an interactive bottom-up and top-down interplay

¹³Note that we used participants' behavioral reports to verify that this manipulation worked as intended, and, on a participant-by-participant basis, we only included those objects in the analysis for which it did.

between perception and cognition (e.g., A. Clark, 2013; Lupyán, 2015). Crucially, however, our study is the first to show that these top-down effects of semantics on perception do not require an extensive learning history to develop. Instead, they can be observed on the very same trial as the visual object is first associated with its semantic meaning (for the N170 component) as well as the very next time when the object is re-encountered, even without the semantic information being present (for the P1 component). Previous research with separate learning and EEG phases, has typically only reported the P1 effect (e.g., Abdel Rahman & Sommer, 2008; Boutonnet & Lupyán, 2015; Maier & Abdel Rahman, 2019; Samaha et al., 2018; Weller et al., 2019), probably because the N170 is short-lived and tied to the specific moment of semantic insight associated with understanding the function of a previously unknown visual object for the first time.

We also conducted an exploratory time-frequency analysis of the same data to check for any effects of semantic information on event-related power in different frequency bands. Unlike ERPs, these changes in event-related power do not need to be tightly phase-locked and time-locked to stimulus onset, which seems plausible given that participants might have taken different amounts of time to understand the functions of different objects. Since we did not have any *a priori* hypothesis about the specific time window, frequency range, and EEG channels at which semantically informed perception would affect event-related power, we used mass-univariate, cluster-based permutation tests (Maris & Oostenveld, 2007; Sassenhagen & Draschkow, 2019) to test for any differences in exploratory fashion. We found no significant effects in the first phase, before any semantic information was presented (as would be expected), but one significant cluster each in the second and third phase. In both cases, there was a statistically significant reduction of event-related power relatively late after stimulus onset (from approximately 600 ms onwards) in the alpha and lower beta frequency bands (approximately 8–20 Hz). Although we did not hypothesize this specific finding, and therefore would like to see it replicated in future studies, it is consistent with previous findings that reduced alpha/beta power facilitates visual-semantic memory formation (Griffiths et al., 2019; Hanslmayr et al., 2009, 2012), presumably because oscillatory activity in this frequency range acts as noise that can hamper the encoding of stimulus-specific information.

Taken together, our study shows that adults are quickly able to associate novel semantic information with previously unfamiliar visual objects, and that this newly acquired semantic information can affect early stages of stimulus processing, typically associated with visual perception itself. Importantly, this top-down effect of semantics on visual perception can be observed immediately as the understanding of the object is being acquired and manifests itself in enlarged amplitudes of ERP components associated with lower-level visual perception (P1 component) and higher-level visual perception (N170 component), as well as in reduced ERP amplitudes in an ERP component associated with semantic precessing demands (N400 components) and reduced alpha/beta band power. This speaks against a modular view of visual perception being encapsulated from higher-level cognitive functions (e.g., Firestone & Scholl, 2016; Fodor, 1983) and favors theories of vision as an interactive, context-sensitive, and prediction-driven process (e.g., Ahissar & Hochstein, 2004; Friston & Kiebel, 2009; Lupyán et al., 2020).

2.3.1 Bonus: The hu-neuro-pipeline package. In this section, I would like to introduce one further research output that, although it is not (yet) captured in a written publication, has already had some impact in my own research bubble and that has taught me a lot while working on it. For the EEG analyses presented in the Enge et al. (2023) paper, I wrote my custom EEG analysis pipeline, based on a previously published pipeline from our lab (see Frömer et al., 2018 and their code and data published at <https://osf.io/hdxvb>). The rationale of the original analysis pipeline was to carry out a number of standard EEG preprocessing steps, including re-referencing, correction of eye artifacts, frequency-domain filtering, epoching, and the rejection of bad epochs based on amplitude thresholds. Unlike in traditional EEG processing workflows (e.g., Luck, 2014), the preprocessed epochs are than *not* averaged into evoked potentials, which would be suitable for statistical analyses using traditional repeated-measures tests from the general linear model (GLM) family, such as a paired *t*-test or a repeated-measures analysis of variance (rmANOVA). Instead, the EEG data remain at the single trial level and are entered directly into a linear mixed-effects model (LMM), which can adequately handle the repeated-measures structure of the data

(with trials nested in participants and/or stimuli) via an appropriately specified random effects structure (Bürki et al., 2018; Frömer et al., 2018; Kretzschmar & Alday, 2023; Volpert-Esmond et al., 2021). Compared to averaging and using a GLM, modeling the single trial data with LMMs provides a number of advantages, including:

- the ability to include both participants *and* stimuli as random effects, which is necessary to maintain adequate false-positive error control (Bürki et al., 2018; Judd et al., 2012) and to allow for inference from the specific sample of stimuli to the larger population from which they were drawn (H. H. Clark, 1973; Yarkoni, 2020),
- the ability to include trial- and stimulus-level covariates (e.g., stimulus characteristics, fatigue, drift; Volpert-Esmond et al., 2021),
- the ability to include not just categorical predictors (i.e., factorial manipulations), but also continuous predictors (e.g., stimulus ratings, parametric manipulations; Brown, 2021; Frömer et al., 2018),
- the ability to include person-level predictors (e.g., age, gender, test scores), and
- the ability to handle unbalanced designs, that is, an uneven number of trials per participant and conditions, which is inevitable in some experimental designs (e.g., Enge et al., 2023; Fröber et al., 2017) as well as when conditions and/or participants differ in their number of epochs after artifact rejection (which will almost always be the case).

The Frömer et al. (2018) pipeline implemented this procedure of standard EEG preprocessing plus single trial LMMs as a collection of scripts in the MATLAB language (The MathWorks Inc., Natick, Massachusetts) using the EEGLAB toolbox (Delorme & Makeig, 2004). However, a few disadvantages of this implementations are that (a) MATLAB is a commercial software, which costs up to \$1000 per user and is therefore not readily usable outside of well-funded research institutions, (b) psychologists are typically not trained in using the MATLAB language, and (c) the code for the pipeline was not structured, documented, and version-controlled as is recommended based on current best practices in research software development (e.g., Barker et al., 2022; Scheliga et al., 2019). In the re-implementation which I created for the Enge et al. (2023) paper and subsequently published as a standalone package (available at <https://github.com/alexenge/hu-neuro-pipeline>), I tried to overcome these limitations and make some further improvements by:

- (1) using a widely used open-source (Python) instead of proprietary (MATLAB) software language,
- (2) building the pipeline on top of MNE-Python (Gramfort et al., 2013), a state-of-the-art M/EEG analysis toolbox with a large developer and user base that follows best practices from professional software development,
- (3) creating an R interface, so that psychologists without prior training in Python or MATLAB can use the pipeline out of the box,
- (4) adding support for time-frequency analysis (including single trial analysis and cluster-based permutation tests),
- (5) releasing the code as a package (using the Python Package Index, see <https://pypi.org/project/hu-neuro-pipeline>), instead of as a collection scripts, so that it can be installed and imported more easily,
- (6) releasing the code under version control, so that users can examine and install older versions of the package,

- (7) adding a publicly accessible documentation website (see <https://hu-neuro-pipeline.readthedocs.io>) that explains the installation, usage, and processing details of the pipeline, both for Python and R users, and
- (8) creating publicly accessible course materials on the pipeline package itself (see <https://github.com/alexenge/hu-neuro-pipeline-workshop>), time frequency analysis (see <https://github.com/alexenge/tfr-workshop>), and EEG analysis in general (see <https://alexenge.github.io/intro-to-eeg>).

The pipeline is designed with a user-interface that consists of one high-level function, `pipeline.group_pipeline`, that can be used to process the raw data from an EEG group study up to the point where the data are ready for single trial LMM analysis. More precisely, it reads the raw data from each participant, optionally resamples it to a lower sampling rate, optionally interpolates any bad channels (which can also be automatically detected), re-references the data (per default to an average reference), optionally performs eye artifact correction (using the semi-automatic multiple source eye correction procedure or fully automatic independent component analysis), filters the data in the frequency domain (per default between 0.1 and 40 Hz), segments the continuous data into epochs based on event markers, reads any accompanying behavioral-experimental log files and matches them to the corresponding EEG epochs, rejects “bad” epochs based on a peak-to-peak amplitude threshold, and computes the single trial amplitudes for any EEG components of interest by averaging across their *a priori* defined time windows and regions of interest. At the group level, the pipeline combines the single trial amplitudes from all participants into one large data frame that can be used as-is for LMM analysis, e.g., using the `lmer` function from the `lme4` package in R (Bates et al., 2015). Additionally, the pipeline also outputs evoked potentials (i.e., averaged waveforms for all participants, experimental conditions, and channels) that can be used for visualization or to compare the statistical results from LMMs and *t*-tests or rmANOVAs, as well as a metadata file that can be referred to when wanting to check the parameters and software versions that were used when running the pipeline.

In sum, the `hu-neuro-pipeline` package provides a relatively user friendly and state of the art EEG analysis pipeline that can be used for LMM analysis of single trial event-related potentials and time-frequency data. Any questions, problems, or improvements to the package can be suggested via the issue tracking system on GitHub (<https://github.com/alexenge/hu-neuro-pipeline/issues>).

3 General discussion

In this thesis, I presented three empirical studies on the development of semantic cognition and its connections to the visual system. In the first study, we conducted a meta-analysis of fMRI studies to identify the brain areas involved in semantic cognition in children and their similarity or dissimilarity to the adult semantic system. In the second study, we tested how learning to read during childhood, that is, learning to associate previously meaningless visual shapes with phonological and semantic information, changes amplitudes and patterns of brain activity in response to written and spoken words. In the third study, we tested if learning semantic information about previously unfamiliar visual objects changes early perception-related brain activity in response to these objects in adults. As mentioned in the introduction, these three studies had not been designed to strictly build up on each other and they are rather diverse in terms of their participant population, study design, and methods used to probe brain activity. Nevertheless, in this discussion I will briefly lay out some connections between our main findings and how they may or may not inform future research as well as everyday life applications.

3.1 Visual-semantic processing in reading and object recognition

The interactions between the visual system and the semantic system of the human brain are clearly experience dependent: From the moment we are born, we are bombarded with visual experiences, but it is only over time that we learn to understand what they mean. We learn which faces are important to remember and what to expect from them, what can be done with different kinds of visual objects, and, eventually, how to extract meaning from written letters and other visual symbols. The visual system itself of course also develops further after birth (for review, see, e.g., Murphy & Monteiro, 2024; Siu & Murphy, 2018), and part of this may be driven by non-visual influences (e.g., from language and semantics; see, e.g., Clarke et al., 2016; Dehaene et al., 2015; Lupyan et al., 2020). However, compared to the visual system, our semantics is much more dependent on our local environment, language, and culture, and differs more strongly from our evolutionary cousins. The semantic system and the visual-semantic connections are therefore likely to change a lot more radically over the course of human development. Given this hypothesis, it was surprising to find very strong convergence between the adult semantic system and the semantic system in children in our meta-analysis of fMRI studies (see Section 2.1 and Enge et al., 2021). We found that virtually all regions active during semantic processing in adults are already active in children, except potentially the more anterior parts of the left temporal lobe (ATL). The ATL is often viewed as an amodal semantic “hub” that dynamically integrates the specific, modal information from the sensory and motor systems (e.g., Chiou & Lambon Ralph, 2019; Lambon Ralph et al., 2017; Patterson et al., 2007). The fact that its activity level does not yet seem to be fully developed in the groups of children included in our meta-analysis suggests that this region, unlike other parts of the semantic network, develops relatively late and might continue to do so throughout adolescence.

However, one should also be cautious not to over-interpret any group difference or lack thereof in our meta-analysis, for a number of reasons. First, the tasks used to assess semantic processing in adults and children often differ from each other in systematic ways (e.g., it is more typical to use tasks with written rather than spoken materials in adults as compared to children; for an extended discussion of semantic task effects in meta-analyses, see Binder et al., 2009). These systematically different tasks might then lead to systematically different results at the meta-analytic level. Second, the activation likelihood estimation (ALE) procedure used in our and most other MRI-based meta-analyses only uses the *location* of reported peaks of fMRI activation from the original studies, not their *activation strength* (i.e., effect size). The meta-analytic maps then show where there is statistically significant overlap of reported fMRI activations regardless of their strength, meaning that a very weakly but consistently activated region might show up in the meta-analysis, whereas a strongly active region with a more variable location across participants or studies might not. Any group differences between children and adults therefore reflect group differences in consistency or reliability of activation, not necessarily in activation strength. Third, regions may differ

in their internal processing (e.g., which algorithms are run or how quickly information is processed) in ways that cannot be picked up using fMRI, which means that they will be missed by any individual fMRI study as well as by our meta-analysis. Finally, the studies included in our meta-analysis mostly tested older children and young adolescents (grand mean age = 10.1 years, range of mean ages = 5.5–12.8 years, total age range 4–15 years). This may be the case because older children are easier to recruit and more compliant inside the MRI scanner, therefore making it easier to obtain sufficient high-quality data. Once a greater number fMRI studies on semantic processing in younger children become available, it would be worthwhile to extend our meta-analysis to check if the semantic system in younger children differs from that in older children and adults.

In the meta-analysis, we identified brain regions that are involved in semantic processing, at least to the degree that can be measured using BOLD fMRI. We found that this set of regions is largely similar to that observed in adults (e.g., Binder et al., 2009; Jackson, 2021; Rodd et al., 2015). This of course does not speak to the question of *what* these regions are doing computationally and *how* they contribute to making sense of visual or other sensory information. To that end, I also presented two original research studies that focused on two example cases of learning visual-semantic information, namely children's learning to read and adults' learning the function of unfamiliar visual objects.

In the study on learning to read, we used fMRI to track longitudinal changes in BOLD activity amplitude and BOLD activity patterns in response to written words. Children received reading instruction for approximately 1.5 years, during which they learnt to associate written letters (in the Hindi/Devanagari script) with speech sounds and word meaning. The design of our fMRI experiment included a comparison between real words, which carry both phonological and semantic information, and pseudowords, which carry the same amount of phonological information but do not carry any semantic information. This contrast between words and pseudowords, which is often used to isolate semantic processing (Binder et al., 2009; Enge et al., 2021), did not elicit any reliable BOLD activation in any brain area. This is in contrast with our meta-analytic results, which showed a reliable network of regions for this type of task, including the left MTG, left IFG, and bilateral dmPFC/pre-SMA. It is worth noting that we presented all of our stimuli in two modalities (auditory/spoken and visual/written). The lack of semantic activation for *written* words versus pseudowords may be explained by the fact that within the 16 months of reading intervention, participants might not have become proficient enough to read and understand most of the words in the short period of time for which they were presented. However, we also did not observe any semantic activation for *spoken* words versus pseudowords, even though children at this age have typically fully developed spoken language comprehension skills (see, e.g., Skeide & Friederici, 2016). Therefore, these null effects might rather be driven by a lack of statistical power or low data quality in our fMRI dataset.

While we did not obtain any evidence for purely semantic activation (as defined by the words versus pseudowords contrast), we did find that learning to read led to changes in BOLD activity for other contrasts, including written words versus false fonts. Like pseudowords, false fonts are visually similar to written words, but they do not contain any phonological or semantic information, as their constituent "letters" do not exist in the writing system and are therefore not associated with any speech sounds or word meanings. The longitudinal changes in BOLD activity amplitude that we observed for this contrast therefore may reflect phonological processing, semantic processing, or a mixture of the two. For the same contrast, we also observed that BOLD activity patterns in response to spoken and written stimuli in the left ventral occipito-temporal (vOT) cortex became more similar over the course of the study. This may indicate that, as children are learning to read, the vOT starts to respond to spoken and written words in a similar fashion. This may be either because the vOT is already responsive to spoken words since early childhood, and learns to access similar information from written words, or because it develops sensitivity to spoken and written words from scratch during reading acquisition. Either way, it is again difficult to pinpoint if the vOT is engaged in phonological processing, semantic processing, or both, as this contrast (written words versus false fonts) may capture both kinds of processes.

In our third study, we investigated a different example case of learning visual-semantic information, namely learning the function of previously unfamiliar real-world objects. We found that discovering the function of an object, as compared to viewing it without knowing its function, affected cortical processing within the first 200 ms after seeing the object, indicating that semantic information influences even relatively early stages of visual perception. Of note, these top-down effects of semantics on visual perception could be observed immediately, that is, on the very same trial that the information has been learned as well as one trial later. This extends previous findings on top-down effects which used an extensive learning period before testing for any learning-related changes (e.g., Abdel Rahman & Sommer, 2008; Boutonnet & Luyan, 2015; Maier & Abdel Rahman, 2019; Samaha et al., 2018; Weller et al., 2019).

The cognitive functions of reading and visual object recognition are highly similar in many ways: Both rely on visual input from the retina that is projected via the LGN in the thalamus to early visual cortex and from there along the ventral visual stream until the word or object is recognized (see Figure 1A in the Introduction). Higher-level visual regions in the ventral stream, located on the lateral and ventral side of the occipito-temporal cortex, then contain clusters that respond preferentially to specific categories, including visual words (the visual word form area; VWFA) and visual objects (the lateral occipital complex; LO; see Figure 1B in the Introduction). This led to the proposal of the neural recycling hypothesis, namely that during learning to read, the brain learns to “reuse” patches of visual cortex that had originally evolved for recognizing other types of visual stimuli (e.g., faces or body parts; Cohen et al., 2000; Dehaene & Cohen, 2007, 2011; Dehaene-Lambertz et al., 2018; Kubota et al., 2023; Nordt et al., 2021). Furthermore, both reading and visual object recognition are compositional, in that smaller visual units (e.g., individual letters of a word or the characteristic visual features of a face) need to be arranged in a certain way such that our brain is able to make sense of the stimulus as a whole.

Nevertheless, there are also a number of important differences between reading and visual object recognition: First, reading and writing began relatively late in our evolutionary history (approximately 5000–7000 years ago; e.g., Houston, 2004), whereas visual object recognition has been a core function of the primate brain for a much longer period of time, likely leading to more specialized cortical circuitry. This explains why learning to read is relatively effortful and requires explicit instruction at a relatively late age, whereas recognizing different types of objects and learning their function occurs relatively effortlessly from the first months of infancy onward. Second, the correlations between visual and semantic information are different: In reading, visual shapes (typically straight and curved black line segments on a white background) are arbitrarily related to the meaning of the word,¹⁴ whereas in visual object recognition, visual shapes and features are highly diagnostic of the category and function of an object. Finally, reading—at least initially—requires access to the language network in order to recode the written letters into phonemes and assemble the phonemes into an auditory word form, the meaning of which can then be retrieved from the mental lexicon. This does not seem necessary in order to understand most visual objects and act appropriately on them. However, more proficient readers rely less on this process of phonological recoding and are able to read visual words as a whole, likely treating them more similar to other visual objects that one has acquired expertise with through repeated presentation and through a reliable association between visual and semantic information.

In our EEG study on object recognition, we found that learning semantic information, namely the function of the object, altered event-related potentials that are associated with visual perception of the object. Similar top-down effects of semantic knowledge (e.g., Abdel Rahman & Sommer, 2008; Maier & Abdel Rahman, 2019; Samaha et al., 2018; Weller et al., 2019) and other kinds of non-visual information (e.g., language and emotion; Boutonnet & Luyan, 2015; Eiserbeck et al., 2024; Luyan et al., 2020; Phelps et al., 2016; Thierry, 2016) are well established and speak for a tight interplay between visual and higher-level cognitive functions (e.g., Ahissar & Hochstein, 2004; Churchland et al., 1994; A. Clark, 2013; Luyan, 2015). In reading research, on the other hand, the contrast between pseudowords and words is often used

¹⁴For example, the words *cat* and *car* are visually more similar than *cat* and *dog*, even though the latter two are much more closely related semantically.

to isolate semantic processing, but it remains unknown if similar top-down effects exist, that is, if the visual perception of a word changes depending on knowing versus not knowing its meaning.

3.2 Implications for future research

Each of the present studies as well as their intersections provide ample room for additional research. Some of this research would already be possible today, whereas other projects would depend on the development of better data acquisition methods, computational models, and substantive theories (as elaborated on in Section 3.4 below).

Our meta-analysis of fMRI studies on semantic cognition in children provides an overview of the developing semantic system based on the current state of the literature. Although already based on a relatively large number of experiments ($N = 50$), adding even more experiments as they get published would definitely make the analysis stronger, as it would reduce the false positive and false negative error rate (see, e.g., Eickhoff et al., 2016). This is especially true given that older fMRI experiments (average year of publication in our meta-analysis = 2010) tend to have smaller sample sizes (Szucs & Ioannidis, 2020) and therefore lead to greater uncertainty in the meta-analysis (Eickhoff et al., 2009; Eickhoff et al., 2012; Turkeltaub et al., 2002). Including additional (future) studies would also help to conduct more meaningful and sensitive sub-analyses, including the differences between different semantic task types and especially the difference between different age groups, both of which we attempted but were limited by the relatively low sample sizes of the sub-groups. Additionally, our classification of experiments into three different categories of semantic tasks (semantic world knowledge, semantic relatedness judgments, and object categorization) was created *ad hoc* based on the currently available studies. A more theory- and/or data-driven taxonomy of task types with finer-grained categories would certainly enhance the meaningfulness of a future meta-analysis (see Poldrack et al., 2011; Poldrack & Yarkoni, 2016 for an attempt to classify cognitive tasks and concepts). Finally, we hope that with a growing number of shared datasets (e.g., Markiewicz et al., 2021) and statistical brain maps (e.g., Gorgolewski et al., 2015), future meta-analyses can be conducted in an image-based fashion, thereby taking into account the full whole-brain information from the original studies and overcoming the limitations of ALE and other coordinate-based methods.

Our fMRI study on learning to read could be improved and built up upon in many different ways, most of which are mentioned in the Limitations section of the original paper. This includes a larger sample size, a longer and more structured reading intervention, and an event-related fMRI task design that reduces head motion and allows for stimulus-level analyses (e.g., representational similarity analysis). Two of the main novelties of our study were (a) probing the influence of learning to read on phonological and semantic processes, not just visual processes (as in previous longitudinal studies; e.g., Dehaene-Lambertz et al., 2018), and (b) studying learning to read in a socio-economic setting and writing system that has traditionally been neglected in developmental cognitive neuroscience (Hindi/Devanagari). We believe that both of these characteristics should be carried forward in future longitudinal studies, while at the same time keeping in mind the high risk associated with longitudinal studies in general and with cross-cultural work in particular. Ideally, future studies would even try to directly compare learning to read and its neural correlates in different writing systems (e.g., alphabetic versus syllabic versus logographic), to test the degree to which the brain uses similar or dissimilar brain areas for mapping graphemes to phonemes and word meaning. To reach adequate levels of statistical power and guard against false-positive and false-negative findings, future studies should try to obtain a larger sample size, both in terms of the number of participants and in terms of the number of time points per participant, since both of these variables will determine the likelihood of accurately capturing change over time. Should it turn out to be practically infeasible to obtain both, it might be worthwhile to test either a large number of children a few times (e.g., 100 children at baseline, 1 year, and 2 years of learning to read) or a small number of children many times (e.g., 3–5 children every two weeks). The former dataset could be analyzed using group statistics, as we did in the present study (but with much larger statistical power), whereas the

latter dataset could be analyzed using single-participant statistics, as is common practice, e.g., in the non-human primate literature (e.g., Fries & Maris, 2022; Ince et al., 2021; Schwarzkopf & Huang, 2024). Regarding the experimental paradigm, future studies could try to move beyond the artificial conditions of single word reading and implement more natural reading tasks. In adults, it has recently been shown that naturalistic reading of stories activates semantic representations across large areas of association cortex, and that this semantic space is shared with the semantic space activated during listening to stories (Deniz et al., 2019). It would be interesting to track the development of this semantic space longitudinally, both during learning spoken language comprehension in early childhood as well as during learning to read and becoming a proficient reader in late childhood and adolescence.

Our third study demonstrated top-down effects of semantic knowledge on visual object perception but leaves open the mechanisms by which this happens and the brain areas that are involved. Methods with better spatial resolution (e.g., magnetoencephalography; MEG) combined with structural MRI images and source modeling (e.g., Knösche & Haueisen, 2022) or high-field fMRI with layer-specific resolution (e.g., Bandettini et al., 2021; Finn et al., 2021; Huber et al., 2021) could be used to investigate which semantic areas in association cortex influence which perceptual areas along the ventral visual stream. Additionally, and relating this third study back to the developmental aspect of the first and second study, it would be worthwhile to adapt a similar experimental paradigm in children to assess if their learning of everyday objects and their functions (e.g., what to do with a pencil sharpener or bottle opener) shows similar electrophysiological effects as the ones observed in adults learning the function of rare real-world objects in our study. If that turns out to be the case, one may also include other types of visual stimuli for which visual-semantic associations can be learned, including drawings, symbols, and written words, to test if there are general principles for learning visual-semantic information or if this is tied to the specific class of visual stimuli (e.g., words versus natural objects).

3.3 Implications for real world applications

All three studies presented in this thesis are firmly within the realm of basic research and, from my personal point of view, do not imply any real world applications. The first and third study were designed to answer principled research questions (*Do the areas involved in semantic cognition in children differ from those in adults?* and *Can semantic information influence the visual perception of objects?*) and without any direct real-world applicability in mind (see also Mook, 1983). Regarding the second study, in theory, longitudinal studies of learning to read would be able to identify interindividual variation in reading-related brain activity that can serve as markers for reading problems and could be used to develop targeted interventions (e.g., using neurofeedback approaches; Taylor & Martz, 2023). However, this would require a much larger sample size, highly reliable behavioral measures, and very sensitive analysis approaches (e.g., based on machine learning), all of which we were unable to implement in our present longitudinal study.

3.4 A need for better methods, questions, and theories

Our current cognitive-scientific research methods are sufficient to establish how certain aspects of brain activity covary with different kinds of experimental manipulations, e.g., which ERP components are sensitive to semantic information when perceiving a visual object. However, we are still far away from understanding the *mechanisms* by which the brain processes semantic information and how they interact with sensory processing.

In part, this is driven by the limitations of non-invasive measurement techniques available in human cognitive neuroscience. While EEG offers excellent temporal resolution, it is only sensitive to the postsynaptic activity of large groups of cortical pyramidal neurons, and even these cannot be precisely localized due to the inverse problem. MEG, while offering greater spatial specificity and signal-to-noise ratio, still suffers from the inverse problem and only picks up magnetic fields from neuronal populations close to

the skull and with a certain spatial orientation. Finally, fMRI relies on blood flow as an indirect proxy of neuronal activity and therefore lacks the temporal resolution to track information processing in real time, while offering a spatial resolution that is much greater than EEG and MEG but still on the order of multiple hundreds of thousands to millions of neurons. Invasive recordings (e.g., electrocorticography with grid or depth electrodes) can directly measure local field potentials and therefore offer the best temporal and spatial resolution, but their use is restricted to certain special cases (e.g., presurgical epilepsy patients) and a very limited number of recording sites. It is therefore unsurprising that our knowledge about higher-level cognitive functions such as language and semantic processing is not as well developed as, for instance, our knowledge about the visual system, which is somewhat easier to study invasively using animal models (Felleman & Van Essen, 1991; Van Essen, 2003).

Three additional obstacles that seem to hamper scientific progress not only in research on visual-semantic processing but in all of cognitive neuroscience are (a) a general lack of theory, (b) the low replicability of research findings, and (c) their limited generalizability.

First, most studies in cognitive neuroscience—and psychological science more broadly—start out with relatively weak predictions that are derived *ad hoc* based on heuristics or in analogy to previous studies, instead of deriving predictions from formal theories (Cummins, 2000; van Rooij & Baggio, 2021; van Rooij & Blokpoel, 2020). Without formal theories or explicit computational models, and with the limitations of non-invasive brain recordings mentioned above, the best one can usually do is to predict that a certain experimental manipulation will affect brain activity to some unknown degree.¹⁵ Combined with the pitfalls of null hypothesis significance testing and the “crud factor”¹⁶ (e.g., Orben & Lakens, 2020), as well as publication bias that favors the publication of “positive” (i.e., statistically significant) effects, this leads to a literature of many individual “islands” of knowledge that are difficult to integrate into a bigger picture across labs, experimental paradigms, and imaging modalities (see also Forscher, 1963; Meehl, 1978).

Second, the last 10–15 years have shown that in the field psychology, results from individual studies cannot be taken at face value, since a substantial proportion of effects (typically $> 50\%$) does not replicate when re-running the same study with a similar or larger sample size (Camerer et al., 2018; Open Science Collaboration, 2015). This is due to a variety of reasons including small sample sizes (Button et al., 2013; Ioannidis, 2005), publication bias (Kühberger et al., 2014; Rosenthal, 1979), and questionable research practices (Kerr, 1998; Simmons et al., 2011). Unfortunately, there have not been any attempts at systematically testing the replicability of findings in cognitive neuroscience (but see Pavlov et al., 2021 for an ongoing attempt in EEG research). This is worrying given that (a) cognitive neuroscience is a directly neighbouring discipline of psychology, (b) sample sizes are even lower in cognitive neuroscience than in psychology (Button et al., 2013; Szucs & Ioannidis, 2020; Szucs & Ioannidis, 2017), likely due to higher costs and efforts during data acquisition, and (c) analysis pipelines in cognitive neuroscience are typically highly complex, leading to highly variable results even when the exact same dataset is analyzed by different teams of researchers (Botvinik-Nezer et al., 2020; Yücel et al., 2024). It would therefore be an important step to make direct replication of EEG and fMRI studies a common practice, in addition to improving the quality of original studies, e.g., by pre-registering study hypotheses and analysis plans (Nosek et al., 2018; Paul et al., 2021; Peikert et al., 2023), by explicating and testing hidden assumptions about tasks and measurements (Elliott et al., 2020; Kragel et al., 2021; Noble et al., 2019; Scheel et al., 2021), by increasing sample sizes (Marek et al., 2022; Szucs & Ioannidis, 2017), and by sharing analysis code and data (Gorgolewski et al., 2016; Markiewicz et al., 2021; Wilkinson et al., 2016).

¹⁵Note that in fMRI, one is typically interested in an *increase* in local BOLD activity, whereas in EEG, even the polarity of an effect can oftentimes not be predicted *a priori*, as it depends on many un-measurable influence such as cortical folding. In both cases, however, there are little to no studies that make more risky (and therefore scientifically more interesting) predictions of point estimates or ranges of effect sizes.

¹⁶That is, the notion that in complex systems like the human mind and brain, everything tends to be correlated with everything else to some unknown degree, and that therefore the null hypothesis taken literally will always be false given sufficient statistical power.

Finally, most studies in cognitive neuroscience are conducted at research institutions in the Global North that can afford the relevant financial and technological resources, and with very selective convenience samples of participants (typically highly educated young adults). Especially when studying higher-level cognitive functions, it would be important to show that findings translate to different sets of participants, stimuli, cultures, languages, and socio-economic strata (e.g., Henrich et al., 2010; Share, 2021; Yarkoni, 2020), or the limitations of findings in these regards should be reflected in the titles and texts of research papers.¹⁷

3.5 Conclusion

In this thesis, I provided different perspectives on how the human brain learns to associate visual information with meaning as well as how semantic information may feed back into visual perception. I provided a developmental perspective of the human cortical semantic system, showing that fMRI studies of semantic cognition in children largely engage the same network of regions as known from adults. I then used learning to read as an example where children learn to associate previously meaningless visual shapes (letters) with speech sounds and meaning, showing that this leads to local changes in word-related BOLD activity amplitudes and response patterns. Finally, I showed that adults can rapidly learn the meaning (i.e., function) of previously unknown real-world objects, and that this learning affects early brain responses to these objects, indicating that semantic knowledge can rapidly feed back into visual perception. Even though the present findings do not integrate into an overarching theory of visual-semantic development, I hope that they stimulate future research towards that goal. I also hope that some of the methods developed for these projects such as our implementation of bias assessment for MRI-based meta-analysis, our implementation of whole-brain linear mixed-effects models for longitudinal fMRI analysis, and our single trial EEG preprocessing pipeline can be reused for future projects on these and other topics in the mind and brain sciences.

¹⁷But see Mook (1983) for an argument that generalizable findings are not always a goal in basic research.

References

- Abdel Rahman, R., & Sommer, W. (2008). Seeing what we know and understand: How knowledge shapes perception. *Psychonomic Bulletin & Review*, 15(6), 1055–1063.
<https://doi.org/10.3758/PBR.15.6.1055>
- Acar, F., Seurinck, R., Eickhoff, S. B., & Moerkerke, B. (2018). Assessing robustness against potential publication bias in activation likelihood estimation (ALE) meta-analyses for fMRI. *PLOS ONE*, 13(11), e0208177. <https://doi.org/10.1371/journal.pone.0208177>
- Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, 8(10), 457–464. <https://doi.org/10.1016/j.tics.2004.08.011>
- Albajes-Eizagirre, A., Solanes, A., & Radua, J. (2019). Meta-analysis of non-statistically significant unreported effects. *Statistical Methods in Medical Research*, 28(12), 3741–3754.
<https://doi.org/10.1177/0962280218811349>
- Albajes-Eizagirre, A., Solanes, A., Vieta, E., & Radua, J. (2019). Voxel-based meta-analysis via permutation of subject images (PSI): Theory and implementation for SDM. *NeuroImage*, 186, 174–184. <https://doi.org/10.1016/j.neuroimage.2018.10.077>
- Alizadeh, S., Jamalabadi, H., Schönauer, M., Leibold, C., & Gais, S. (2017). Decoding cognitive concepts from neuroimaging data using multivariate pattern analysis. *NeuroImage*, 159, 449–458.
<https://doi.org/10.1016/j.neuroimage.2017.07.058>
- Athanasiopoulos, P., & Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, 37(5-6), 393–412.
<https://doi.org/10.1080/02643294.2020.1769050>
- Bandettini, P. A., Huber, L., & Finn, E. S. (2021). Challenges and opportunities of mesoscopic brain mapping with fMRI. *Current Opinion in Behavioral Sciences*, 40, 189–200.
<https://doi.org/10.1016/j.cobeha.2021.06.002>
- Barker, M., Chue Hong, N. P., Katz, D. S., Lamprecht, A.-L., Martinez-Ortiz, C., Psomopoulos, F., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., & Honeyman, T. (2022). Introducing the FAIR Principles for research software. *Scientific Data*, 9(1), 622.
<https://doi.org/10.1038/s41597-022-01710-x>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Behrmann, M., & Plaut, D. C. (2020). Hemispheric organization for visual object recognition: A theoretical account and empirical evidence. *Perception*, 49(4), 373–404.
<https://doi.org/10.1177/0301006619899049>
- Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1), 65–98. <https://doi.org/10.1137/141000671>
- Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, 19(12), 2767–2796. <https://doi.org/10.1093/cercor/bhp055>
- Bortolini, T., Laport, M. C., Latgé-Tovar, S., Fischer, R., Zahn, R., de Oliveira-Souza, R., & Moll, J. (2024). The extended neural architecture of human attachment: An fMRI coordinate-based meta-analysis of affiliative studies. *Neuroscience & Biobehavioral Reviews*, 159, 105584.
<https://doi.org/10.1016/j.neubiorev.2024.105584>
- Botvinik-Nezer, R., Holzmeister, F., Camerer, C. F., Dreber, A., Huber, J., Johannesson, M., Kirchler, M., Iwanir, R., Mumford, J. A., Adcock, R. A., Avesani, P., Baczkowski, B. M., Bajracharya, A., Bakst, L., Ball, S., Barilari, M., Bault, N., Beaton, D., Beitner, J., ... Schonberg, T. (2020). Variability in the analysis of a single neuroimaging dataset by many teams. *Nature*, 582(7810), 84–88. <https://doi.org/10.1038/s41586-020-2314-9>
- Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience*, 35(25), 9329–9335. <https://doi.org/10.1523/JNEUROSCI.5111-14.2015>

- Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, 4(1), 1–19. <https://doi.org/10.1177/2515245920960351>
- Bürki, A., Frossard, J., & Renaud, O. (2018). Accounting for stimulus and participant effects in event-related potential analyses to increase the replicability of studies. *Journal of Neuroscience Methods*, 309, 218–227. <https://doi.org/10.1016/j.jneumeth.2018.09.016>
- Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376. <https://doi.org/10.1038/nrn3475>
- Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., ... Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 2(9), 637–644. <https://doi.org/10.1038/s41562-018-0399-z>
- Chen, G., Saad, Z. S., Britton, J. C., Pine, D. S., & Cox, R. W. (2013). Linear mixed-effects modeling approach to fMRI group analysis. *NeuroImage*, 73, 176–190. <https://doi.org/10.1016/j.neuroimage.2013.01.047>
- Chiou, R., & Lambon Ralph, M. A. (2019). Unveiling the dynamic interplay between the hub- and spoke-components of the brain's semantic system and its impact on human behaviour. *NeuroImage*, 199, 114–126. <https://doi.org/10.1016/j.neuroimage.2019.05.059>
- Churchland, P. S., Ramachandran, V. S., & Sejnowski, T. J. (1994). A critique of pure vision. In C. Koch & J. L. Davis (Eds.), *Computational Neuroscience. Large-Scale Neuronal Theories of the Brain* (pp. 23–60). MIT Press.
- Chyl, K., Fraga-González, G., Brem, S., & Jednoróg, K. (2021). Brain dynamics of (a)typical reading development—a review of longitudinal studies. *Npj Science of Learning*, 6(1), 1–9. <https://doi.org/10.1038/s41539-020-00081-5>
- 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. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior*, 12(4), 335–359. [https://doi.org/10.1016/S0022-5371\(73\)80014-3](https://doi.org/10.1016/S0022-5371(73)80014-3)
- Clarke, A., Pell, P. J., Ranganath, C., & Tyler, L. K. (2016). Learning warps object representations in the ventral temporal cortex. *Journal of Cognitive Neuroscience*, 28(7), 1010–1023. https://doi.org/10.1162/jocn_a_00951
- Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., & Michel, F. (2000). The Visual Word Form Area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123(2), 291–307. <https://doi.org/10.1093/brain/123.2.291>
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256. <https://doi.org/10.1037/0033-295x.108.1.204>
- Cui, L., Ye, M., Sun, L., Zhang, S., & He, G. (2022). Common and distinct neural correlates of intertemporal and risky decision-making: Meta-analytical evidence for the dual-system theory. *Neuroscience & Biobehavioral Reviews*, 141, 104851. <https://doi.org/10.1016/j.neubiorev.2022.104851>
- Cummins, R. (2000). "How does it work?" versus "What are the laws?": Two conceptions of psychological explanation. In *Explanation and Cognition* (pp. 117–144). MIT Press.
- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2), 384–398. <https://doi.org/10.1016/j.neuron.2007.10.004>
- Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6), 254–262. <https://doi.org/10.1016/j.tics.2011.04.003>

- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, 16(4), 234–244. <https://doi.org/10.1038/nrn3924>
- Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology*, 16(3), 1–34. <https://doi.org/10.1371/journal.pbio.2004103>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Deniz, F., Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *Journal of Neuroscience*, 39(39), 7722–7736. <https://doi.org/10.1523/JNEUROSCI.0675-19.2019>
- Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *NeuroImage*, 53(1), 1–15. <https://doi.org/10.1016/j.neuroimage.2010.06.010>
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73(3), 415–434. <https://doi.org/10.1016/j.neuron.2012.01.010>
- Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167–188. <https://doi.org/fkh923>
- Eickhoff, S. B., Bzdok, D., Laird, A. R., Kurth, F., & Fox, P. T. (2012). Activation likelihood estimation meta-analysis revisited. *NeuroImage*, 59(3), 2349–2361. <https://doi.org/10.1016/j.neuroimage.2011.09.017>
- Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., & Fox, P. T. (2009). Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human Brain Mapping*, 30(9), 2907–2926. <https://doi.org/10.1002/hbm.20718>
- Eickhoff, S. B., Nichols, T. E., Laird, A. R., Hoffstaedter, F., Amunts, K., Fox, P. T., Bzdok, D., & Eickhoff, C. R. (2016). Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage*, 137, 70–85. <https://doi.org/10.1016/j.neuroimage.2016.04.072>
- Eiserbeck, A., Enge, A., Rabovsky, M., & Abdel Rahman, R. (2024). Distrust before first sight? Examining knowledge- and appearance-based effects of trustworthiness on the visual consciousness of faces. *Consciousness and Cognition*, 117, 103629. <https://doi.org/10.1016/j.concog.2023.103629>
- Elliott, M. L., Knott, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychological Science*, 31(7), 792–806. <https://doi.org/10.1177/0956797620916786>
- Embleton, K. V., Haroon, H. A., Morris, D. M., Ralph, M. A. L., & Parker, G. J. M. (2010). Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes. *Human Brain Mapping*, 31(10), 1570–1587. <https://doi.org/10.1002/hbm.20959>
- Enge, A., Abdel Rahman, R., & Skeide, M. A. (2021). A meta-analysis of fMRI studies of semantic cognition in children. *NeuroImage*, 241, 118436. <https://doi.org/10.1016/j.neuroimage.2021.118436>
- Enge, A., Friederici, A. D., & Skeide, M. A. (2020). A meta-analysis of fMRI studies of language comprehension in children. *NeuroImage*, 215, 116858. <https://doi.org/10.1016/j.neuroimage.2020.116858>
- Enge, A., & Skeide, M. A. (2024). *Tracking the neural correlates of learning to read with dense-sampling fMRI*. Unpublished manuscript available from first author.
- Enge, A., Süß, F., & Abdel Rahman, R. (2023). Instant effects of semantic information on visual perception. *Journal of Neuroscience*, 43(26), 4896–4906. <https://doi.org/10.1523/JNEUROSCI.2038-22.2023>

- Enoch, J., McDonald, L., Jones, L., Jones, P. R., & Crabb, D. P. (2019). Evaluating whether sight is the most valued sense. *JAMA Ophthalmology*, 137(11), 1317.
<https://doi.org/10.1001/jamaophthalmol.2019.3537>
- Epstein, R., Harris, A., Stanley, D., & Kanwisher, N. (1999). The Parahippocampal Place Area: Recognition, navigation, or encoding? *Neuron*, 23(1), 115–125.
[https://doi.org/10.1016/s0896-6273\(00\)80758-8](https://doi.org/10.1016/s0896-6273(00)80758-8)
- Farivar, R. (2009). Dorsal–ventral integration in object recognition. *Brain Research Reviews*, 61(2), 144–153. <https://doi.org/10.1016/j.brainresrev.2009.05.006>
- Fedorenko, E., Ivanova, A. A., & Regev, T. I. (2024). The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews Neuroscience*, 25(5), 289–312.
<https://doi.org/10.1038/s41583-024-00802-4>
- Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1), 1–47. <https://doi.org/10.1093/cercor/1.1.1-a>
- Ferstl, E. C., Neumann, J., Bogler, C., & von Cramon, D. Y. (2008). The extended language network: A meta-analysis of neuroimaging studies on text comprehension. *Human Brain Mapping*, 29(5), 581–593. <https://doi.org/10.1002/hbm.20422>
- Finn, E. S., Huber, L., & Bandettini, P. A. (2021). Higher and deeper: Bringing layer fMRI to association cortex. *Progress in Neurobiology*, 207, 101930.
<https://doi.org/10.1016/j.pneurobio.2020.101930>
- Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for “top-down” effects. *Behavioral and Brain Sciences*, 39, e229.
<https://doi.org/10.1017/S0140525X15000965>
- Fodor, J. A. (1983). *The modularity of mind*. MIT Press.
- Fodor, J. A. (1984). Observation reconsidered. *Philosophy of Science*, 51(1), 23–43.
<https://doi.org/10.1086/289162>
- Forscher, B. K. (1963). Chaos in the brickyard. *Science*, 142(3590), 339–339.
<https://doi.org/10.1126/science.142.3590.339.a>
- Forseth, K. J., Kadipasaoglu, C. M., Conner, C. R., Hickok, G., Knight, R. T., & Tandon, N. (2018). A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain*, 141(7), 2112–2126.
<https://doi.org/10.1093/brain/awy120>
- Friederici, A. D. (2006). The neural basis of language development and its impairment. *Neuron*, 52(6), 941–952. <https://doi.org/10.1016/j.neuron.2006.12.002>
- Friederici, A. D. (2017). *Language in our brain: The origins of a uniquely human capacity*. MIT Press.
<https://doi.org/10.7551/mitpress/11173.001.0001>
- Fries, P., & Maris, E. (2022). What to do if N is two? *Journal of Cognitive Neuroscience*, 34(7), 1114–1118. https://doi.org/10.1162/jocn_a_01857
- Frisby, S. L., Halai, A. D., Cox, C. R., Lambon Ralph, M. A., & Rogers, T. T. (2023). Decoding semantic representations in mind and brain. *Trends in Cognitive Sciences*, 27(3), 258–281.
<https://doi.org/10.1016/j.tics.2022.12.006>
- Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1521), 1211–1221.
<https://doi.org/10.1098/rstb.2008.0300>
- Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, 36(1), 67–81. <https://doi.org/10.1007/BF02648022>
- Fröber, K., Stürmer, B., Frömer, R., & Dreisbach, G. (2017). The role of affective evaluation in conflict adaptation: An LRP study. *Brain and Cognition*, 116, 9–16.
<https://doi.org/10.1016/j.bandc.2017.05.003>
- Frömer, R., Maier, M., & Abdel Rahman, R. (2018). Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. *Frontiers in Neuroscience*, 12, 48.
<https://doi.org/10.3389/fnins.2018.00048>

- Gauthier, I., James, T. W., Curby, K. M., & Tarr, M. J. (2003). The influence of conceptual knowledge on visual discrimination. *Cognitive Neuropsychology*, 20(3-6), 507–523.
<https://doi.org/10.1080/02643290244000275>
- Gelman, A., & Stern, H. (2006). The difference between “significant” and “not significant” is not itself statistically significant. *The American Statistician*, 60(4), 328–331.
<https://doi.org/10.1198/000313006X152649>
- Ghio, M., Cara, C., & Tettamanti, M. (2021). The prenatal brain readiness for speech processing: A review on foetal development of auditory and primordial language networks. *Neuroscience & Biobehavioral Reviews*, 128, 709–719. <https://doi.org/10.1016/j.neubiorev.2021.07.009>
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3–8. <https://doi.org/10.2307/1174772>
- Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615), 171–178.
<https://doi.org/10.1038/nature18933>
- Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S. S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., ... Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, 3(1), 160044. <https://doi.org/10.1038/sdata.2016.44>
- Gorgolewski, K. J., Varoquaux, G., Rivera, G., Schwarz, Y., Ghosh, S. S., Maumet, C., Sochat, V. V., Nichols, T. E., Poldrack, R. A., Poline, J.-B., Yarkoni, T., & Margulies, D. S. (2015). NeuroVault.org: A web-based repository for collecting and sharing unthresholded statistical maps of the human brain. *Frontiers in Neuroinformatics*, 9. <https://doi.org/10.3389/fninf.2015.00008>
- Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 7, 267. <https://doi.org/10.3389/fnins.2013.00267>
- Griffiths, B. J., Mayhew, S. D., Mullinger, K. J., Jorge, J., Charest, I., Wimber, M., & Hanslmayr, S. (2019). Alpha/beta power decreases track the fidelity of stimulus-specific information. *eLife*, 8, e49562. <https://doi.org/10.7554/eLife.49562>
- Hagoort, P. (2016). Chapter 28 - MUC (memory, unification, control): A model on the neurobiology of language beyond single word processing. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 339–347). Academic Press. <https://doi.org/10.1016/B978-0-12-407794-2.00028-6>
- Hanslmayr, S., Spitzer, B., & Bäuml, K.-H. (2009). Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cerebral Cortex*, 19(7), 1631–1640.
<https://doi.org/10.1093/cercor/bhn197>
- Hanslmayr, S., Staudigl, T., & Fellner, M.-C. (2012). Oscillatory power decreases and long-term memory: The information via desynchronization hypothesis. *Frontiers in Human Neuroscience*, 6. <https://doi.org/10.3389/fnhum.2012.00074>
- Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). *Doing Meta-Analysis with R: A Hands-on Guide* (1st ed.). Chapman & Hall/CRC Press.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *arXiv*. <https://doi.org/10.48550/arXiv.1512.03385>
- Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Wicklin, C. V., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLOS ONE*, 14(10), e0223792. <https://doi.org/10.1371/journal.pone.0223792>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2-3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402. <https://doi.org/10.1038/nrn2113>

- Houston, S. D. (Ed.). (2004). *The first writing: Script invention as history and process*. Cambridge University Press.
- Hsieh, P.-J., Vul, E., & Kanwisher, N. (2010). Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex. *Journal of Neurophysiology*, 103(3), 1501–1507.
<https://doi.org/10.1152/jn.00812.2009>
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160(1), 106.
<https://doi.org/10.1113/jphysiol.1962.sp006837>
- Huber, L. R., Poser, B. A., Bandettini, P. A., Arora, K., Wagstyl, K., Cho, S., Goense, J., Nothnagel, N., Morgan, A. T., van den Hurk, J., Müller, A. K., Reynolds, R. C., Glen, D. R., Goebel, R., & Gulban, O. F. (2021). LayNii: A software suite for layer-fMRI. *NeuroImage*, 237, 118091.
<https://doi.org/10.1016/j.neuroimage.2021.118091>
- Hupé, J. M., James, A. C., Payne, B. R., Lomber, S. G., Girard, P., & Bullier, J. (1998). Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature*, 394(6695), 784–787. <https://doi.org/10.1038/29537>
- Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532(7600), 453–458.
<https://doi.org/10.1038/nature17637>
- Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, 76(6), 1210–1224. <https://doi.org/10.1016/j.neuron.2012.10.014>
- Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of population prevalence. *eLife*, 10, e62461. <https://doi.org/10.7554/eLife.62461>
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), e124.
<https://doi.org/10.1371/journal.pmed.0020124>
- Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, 40(3), 366–371.
<https://doi.org/10.1509/jmkr.40.3.366.19237>
- Jackson, R. L. (2021). The neural correlates of semantic control revisited. *NeuroImage*, 224, 117444.
<https://doi.org/10.1016/j.neuroimage.2020.117444>
- Jackson, R. L., Lambon Ralph, M. A., & Pobric, G. (2015). The timing of anterior temporal lobe involvement in semantic processing. *Journal of Cognitive Neuroscience*, 27(7), 1388–1396.
https://doi.org/10.1162/jocn_a_00788
- Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, 103(1), 54–69. <https://doi.org/10.1037/a0028347>
- Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17(11), 4302.
<https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997>
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2(3), 196–217. https://doi.org/10.1207/s15327957pspr0203_4
- Knösche, T. R., & Haueisen, J. (2022). *EEG/MEG source reconstruction: Textbook for electro- and magnetoencephalography*. Springer.
- Kragel, P. A., Han, X., Kraynak, T. E., Gianaros, P. J., & Wager, T. D. (2021). Functional MRI can be highly reliable, but it depends on what you measure: A commentary on Elliott et al. (2020). *Psychological Science*, 32(4), 622–626. <https://doi.org/10.1177/0956797621989730>
- Kretzschmar, F., & Alday, P. M. (2023). Principles of statistical analyses: Old and new tools. In M. Grimaldi, E. Brattico, & Y. Shtyrov (Eds.), *Language electrified. Techniques, methods, applications, and future perspectives in the neurophysiological investigation of language*. Humana.
<https://doi.org/10.31234/osf.io/nyj3k>. <https://doi.org/10.31234/osf.io/nyj3k>

- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
<https://doi.org/10.1145/3065386>
- Kubota, E., Grill-Spector, K., & Nordt, M. (2023). Rethinking cortical recycling in ventral temporal cortex. *Trends in Cognitive Sciences*. <https://doi.org/10.1016/j.tics.2023.09.006>
- Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLoS ONE*, 9(9), e105825.
<https://doi.org/10.1371/journal.pone.0105825>
- Lambon Ralph, M. A., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience*, 18(1), 42–55.
<https://doi.org/10.1038/nrn.2016.150>
- Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, 23(11), 571–579.
[https://doi.org/10.1016/S0166-2236\(00\)01657-X](https://doi.org/10.1016/S0166-2236(00)01657-X)
- Lamme, V. A. F., Supèr, H., & Spekreijse, H. (1998). Feedforward, horizontal, and feedback processing in the visual cortex. *Current Opinion in Neurobiology*, 8(4), 529–535.
[https://doi.org/10.1016/S0959-4388\(98\)80042-1](https://doi.org/10.1016/S0959-4388(98)80042-1)
- Lewis, M., & Mitchell, M. (2024). Using counterfactual tasks to evaluate the generality of analogical reasoning in large language models. *arXiv*. <https://doi.org/10.48550/arXiv.2402.08955>
- Liu, T. T. (2016). Noise contributions to the fMRI signal: An overview. *NeuroImage*, 143, 141–151.
<https://doi.org/10.1016/j.neuroimage.2016.09.008>
- Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.
- Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of Philosophy and Psychology*, 6(4), 547–569.
<https://doi.org/10.1007/s13164-015-0253-4>
- Lupyan, G., Abdel Rahman, R., 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>
- Machery, E. (2015). Cognitive penetrability: A no-progress report. In J. Zeimbekis & A. Raftopoulos (Eds.), *The Cognitive Penetrability of Perception: New Philosophical Perspectives*. Oxford University Press.
- Madhyastha, T., Peverill, M., Koh, N., McCabe, C., Flournoy, J., Mills, K., King, K., Pfeifer, J., & McLaughlin, K. A. (2018). Current methods and limitations for longitudinal fMRI analysis across development. *Developmental Cognitive Neuroscience*, 33, 118–128. <https://doi.org/gdz7rh>
- Maier, M., & Abdel Rahman, R. (2019). No matter how: Top-down effects of verbal and semantic category knowledge on early visual perception. *Cognitive, Affective, & Behavioral Neuroscience*, 19(4), 859–876. <https://doi.org/10.3758/s13415-018-00679-8>
- Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczko, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., ... Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603(7902), 654–660. <https://doi.org/10.1038/s41586-022-04492-9>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Markiewicz, C. J., Gorgolewski, K. J., Feingold, F., Blair, R., Halchenko, Y. O., Miller, E., Hardcastle, N., Wexler, J., Esteban, O., Goncalves, M., Jwa, A., & Poldrack, R. (2021). The OpenNeuro resource for sharing of neuroscience data. *eLife*, 10, e71774. <https://doi.org/10.7554/eLife.71774>
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. W.H. Freeman.

- McClelland, G. H., Lynch, J. G., Irwin, J. R., Spiller, S. A., & Fitzsimons, G. J. (2015). Median splits, Type II errors, and false-positive consumer psychology: Don't fight the power. *Journal of Consumer Psychology, 25*(4), 679–689. <https://doi.org/10.1016/j.jcps.2015.05.006>
- Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46*(4), 806–834. <https://doi.org/10.1037/0022-006X.46.4.806>
- Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural Brain Research, 6*(1), 57–77. [https://doi.org/10.1016/0166-4328\(82\)90081-X](https://doi.org/10.1016/0166-4328(82)90081-X)
- Mook, D. G. (1983). In defense of external invalidity. *American Psychologist, 38*(4), 379–387. <https://doi.org/10.1037/0003-066X.38.4.379>
- Mooney, C. M. (1957). Age in the development of closure ability in children. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 11*(4), 219–226. <https://doi.org/10.1037/h0083717>
- Murphy, K. M., & Monteiro, L. (2024). Anatomical and molecular development of the human primary visual cortex. *Frontiers in Cellular Neuroscience, 18*. <https://doi.org/10.3389/fncel.2024.1427515>
- Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. *Nature Neuroscience, 14*(9), 1105–1107. <https://doi.org/10.1038/nn.2886>
- Noble, S., Scheinost, D., & Constable, R. T. (2019). A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage, 203*, 116157. <https://doi.org/10.1016/j.neuroimage.2019.116157>
- Nordt, M., Gomez, J., Natu, V. S., Rezai, A. A., Finzi, D., Kular, H., & Grill-Spector, K. (2021). Cortical recycling in high-level visual cortex during childhood development. *Nature Human Behaviour, 1*–12. <https://doi.org/10.1038/s41562-021-01141-5>
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences, 115*(11), 2600–2606. <https://doi.org/10.1073/pnas.1708274114>
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716. <https://doi.org/10.1126/science.aac4716>
- Orben, A., & Lakens, D. (2020). Crud (re)defined. *Advances in Methods and Practices in Psychological Science, 3*(2), 238–247. <https://doi.org/10.1177/2515245920917961>
- Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976–987. <https://doi.org/10.1038/nrn2277>
- Paul, M., Govaart, G. H., & Schettino, A. (2021). Making ERP research more transparent: Guidelines for preregistration. *International Journal of Psychophysiology, 164*, 52–63. <https://doi.org/10.1016/j.ijpsycho.2021.02.016>
- Pavlov, Y. G., Adamian, N., Appelhoff, S., Arvaneh, M., Benwell, C. S. Y., Beste, C., Bland, A. R., Bradford, D. E., Bublitzky, F., Busch, N. A., Clayson, P. E., Cruse, D., Czeszumski, A., Dreber, A., Dumas, G., Ehinger, B., Ganis, G., He, X., Hinojosa, J. A., ... Mushtaq, F. (2021). #EEGManyLabs: Investigating the replicability of influential EEG experiments. *Cortex, 144*, 213–229. <https://doi.org/10.1016/j.cortex.2021.03.013>
- Peikert, A., Ernst, M. S., & Brandmaier, A. M. (2023). Why does preregistration increase the persuasiveness of evidence? A Bayesian rationalization. *PsyArXiv*. <https://doi.org/10.31234/osf.io/cs8wb>
- Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114*(2), 273–315. <https://doi.org/10.1037/0033-295X.114.2.273>
- Perry, C., Ziegler, J. C., & Zorzi, M. (2010). Beyond single syllables: Large-scale modeling of reading aloud with the Connectionist Dual Process (CDP++) model. *Cognitive Psychology, 61*(2), 106–151. <https://doi.org/10.1016/j.cogpsych.2010.04.001>

- Perry, C., Ziegler, J. C., & Zorzi, M. (2013). A computational and empirical investigation of graphemes in reading. *Cognitive Science*, 37(5), 800–828. <https://doi.org/10.1111/cogs.12030>
- Phelps, E. A., Ling, S., & Carrasco, M. (2016). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science*, 17(4), 292–299.
- Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., Parker, D. S., Sabb, F. W., & Bilder, R. M. (2011). The cognitive atlas: Toward a knowledge foundation for cognitive neuroscience. *Frontiers in Neuroinformatics*, 5. <https://doi.org/10.3389/fninf.2011.00017>
- Poldrack, R. A., & Yarkoni, T. (2016). From brain maps to cognitive ontologies: Informatics and the search for mental structure. *Annual Review of Psychology*, 67, 587–612. <https://doi.org/10.1146/annurev-psych-122414-033729>
- Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral and Brain Sciences*, 22(3), 341–365. <https://doi.org/10.1017/S0140525X99002022>
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. <https://doi.org/10.1038/4580>
- Reilly, J., Shain, C., Borghesani, V., Kuhnke, P., Vigliocco, G., Peelle, J. E., Mahon, B. Z., Buxbaum, L. J., Majid, A., Brysbaert, M., Borghi, A. M., De Deyne, S., Dove, G., Papeo, L., Pexman, P. M., Poeppel, D., Lupyán, G., Boggio, P., Hickok, G., ... Vinson, D. (2024). What we mean when we say semantic: Toward a multidisciplinary semantic glossary. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-024-02556-7>
- Reining, L. C., & Wallis, T. S. A. (2024). A psychophysical evaluation of techniques for Mooney image generation. *PeerJ*, 12, e18059. <https://doi.org/10.7717/peerj.18059>
- Rice, G. E., Watson, D. M., Hartley, T., & Andrews, T. J. (2014). Low-level image properties of visual objects predict patterns of neural response across category-selective regions of the ventral visual pathway. *Journal of Neuroscience*, 34(26), 8837–8844. <https://doi.org/10.1523/JNEUROSCI.5265-13.2014>
- Rodd, J. M., Vitello, S., Woollams, A. M., & Adank, P. (2015). Localising semantic and syntactic processing in spoken and written language comprehension: An activation likelihood estimation meta-analysis. *Brain and Language*, 141, 89–102. <https://doi.org/10.1016/j.bandl.2014.11.012>
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86(3), 638–641. <https://doi.org/10.1037/0033-2909.86.3.638>
- Salo, T., Yarkoni, T., Nichols, T. E., Poline, J.-B., Bilgel, M., Bottenhorn, K. L., Jarecka, D., Kent, J. D., Kimbler, A., Nielson, D. M., Oudyk, K. M., Peraza, J. A., Pérez, A., Reeders, P. C., Yanes, J. A., & Laird, A. R. (2023). NiMARE: Neuroimaging meta-analysis research environment. *Aperture Neuro*, 3, 1–32. <https://doi.org/10.52294/001c.87681>
- Samaha, J., Boutonnet, B., Postle, B. R., & Lupyán, G. (2018). Effects of meaningfulness on perception: Alpha-band oscillations carry perceptual expectations and influence early visual responses. *Scientific Reports*, 8(1), 6606. <https://doi.org/10.1038/s41598-018-25093-5>
- Samartsidis, P., Montagna, S., Laird, A. R., Fox, P. T., Johnson, T. D., & Nichols, T. E. (2020). Estimating the prevalence of missing experiments in a neuroimaging meta-analysis. *Research Synthesis Methods*, 11(6), 866–883. <https://doi.org/10.1002/jrsm.1448>
- Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, 56(6), e13335. <https://doi.org/10.1111/psyp.13335>
- Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, 28(9), 3095–3114. <https://doi.org/10.1093/cercor/bhx179>

- Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science*, 16(4), 744–755. <https://doi.org/10.1177/1745691620966795>
- Scheliga, K. S., Pampel, H., Konrad, U., Fritzsch, B., Schlauch, T., Nolden, M., zu Castell, W., Finke, A., Hammitzsch, M., Bertuch, O., & Denker, M. (2019). *Dealing with research software: Recommendations for best practices*. <https://doi.org/10.2312/os.helmholtz.003>
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings*. SAGE Publications, Ltd. <https://doi.org/10.4135/9781483398105>
- Schwarzkopf, D. S., & Huang, Z. (2024). A simple statistical framework for small sample studies. *bioRxiv*, 2023.09.19.558509. <https://doi.org/10.1101/2023.09.19.558509>
- Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96(4), 523–568. <https://doi.org/10.1037/0033-295X.96.4.523>
- Shapson-Coe, A., Januszewski, M., Berger, D. R., Pope, A., Wu, Y., Blakely, T., Schalek, R. L., Li, P. H., Wang, S., Maitin-Shepard, J., Karlupia, N., Dorkenwald, S., Sjostedt, E., Leavitt, L., Lee, D., Troidl, J., Collman, F., Bailey, L., Fitzmaurice, A., ... Lichtman, J. W. (2024). A petavoxel fragment of human cerebral cortex reconstructed at nanoscale resolution. *Science*, 384(6696), eadk4858. <https://doi.org/10.1126/science.adk4858>
- Share, D. L. (2021). Is the science of reading just the science of reading English? *Reading Research Quarterly*, 56(S1), 391–402. <https://doi.org/10.1002/rrq.401>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22(11), 1359–1366. <https://doi.org/10.1177/0956797611417632>
- Siu, C. R., & Murphy, K. M. (2018). The development of human visual cortex and clinical implications. *Eye and Brain*, 10, 25. <https://doi.org/10.2147/EB.S130893>
- Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, 17(5), 323–332. <https://doi.org/10.1038/nrn.2016.23>
- Slivac, K., Hervais-Adelman, A., Hagoort, P., & Flecken, M. (2021). Linguistic labels cue biological motion perception and misperception. *Scientific Reports*, 11(1), 17239. <https://doi.org/10.1038/s41598-021-96649-1>
- Sun, J., Zheng, C., Xie, E., Liu, Z., Chu, R., Qiu, J., Xu, J., Ding, M., Li, H., Geng, M., Wu, Y., Wang, W., Chen, J., Yin, Z., Ren, X., Fu, J., He, J., Yuan, W., Liu, Q., ... Li, Z. (2024). A survey of reasoning with foundation models. *arXiv*. <https://doi.org/10.48550/arXiv.2312.11562>
- Szucs, D., & Ioannidis, J. P. (2020). Sample size evolution in neuroimaging research: An evaluation of highly-cited studies (1990-2012) and of latest practices (2017-2018) in high-impact journals. *NeuroImage*, 221, 117164. <https://doi.org/10.1016/j.neuroimage.2020.117164>
- Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, 15(3), e2000797. <https://doi.org/10.1371/journal.pbio.2000797>
- Taylor, S. F., & Martz, M. E. (2023). Real-time fMRI neurofeedback: The promising potential of brain-training technology to advance clinical neuroscience. *Neuropsychopharmacology*, 48(1), 238–239. <https://doi.org/10.1038/s41386-022-01397-z>
- Teufel, C., Dakin, S. C., & Fletcher, P. C. (2018). Prior object-knowledge sharpens properties of early visual feature-detectors. *Scientific Reports*, 8(1), 10853. <https://doi.org/10.1038/s41598-018-28845-5>
- Thierry, G. (2016). Neurolinguistic relativity: How language flexes human perception and cognition. *Language Learning*, 66(3), 690–713. <https://doi.org/10.1111/lang.12186>
- Turkeltaub, P. E., Eden, G. F., Jones, K. M., & Zeffiro, T. A. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*, 16(3), 765–780. <https://doi.org/10.1006/nimg.2002.1131>
- Van Essen, D. C. (2003). Organization of visual areas in macaque and human cerebral cortex. In L. M. Chalupa & J. S. Werner (Eds.), *The Visual Neurosciences* (Vol. 2, pp. 507–521). MIT Press.

- van Rooij, I., & Baggio, G. (2021). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *Perspectives on Psychological Science*, 16(4), 682–697. <https://doi.org/10.1177/1745691620970604>
- van Rooij, I., & Blokpoel, M. (2020). Formalizing verbal theories: A tutorial by dialogue. *Social Psychology*, 51(5), 285–298. <https://doi.org/10.1027/1864-9335/a000428>
- Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, 30(4), 1414–1432. <https://doi.org/10.1016/j.neuroimage.2005.11.002>
- Vin, R., Blauch, N. M., Plaut, D. C., & Behrmann, M. (2024). Visual word processing engages a hierarchical, distributed, and bilateral cortical network. *iScience*, 27(2), 108809. <https://doi.org/10.1016/j.isci.2024.108809>
- Visser, M., Jefferies, E., & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: A meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, 22(6), 1083–1094. <https://doi.org/10.1162/jocn.2009.21309>
- Volkert-Esmond, H. I., Page-Gould, E., & Bartholow, B. D. (2021). Using multilevel models for the analysis of event-related potentials. *International Journal of Psychophysiology*, 162, 145–156. <https://doi.org/10.1016/j.ijpsycho.2021.02.006>
- von Helmholtz, H. (1867). *Handbuch der physiologischen Optik* (Vol. 9). Voss.
- 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
- Wenger, E., Brozzoli, C., Lindenberger, U., & Lövdén, M. (2017). Expansion and renormalization of human brain structure during skill acquisition. *Trends in Cognitive Sciences*, 21(12), 930–939. <https://doi.org/10.1016/j.tics.2017.09.008>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Williams, M. A., Baker, C. I., Op de Beeck, H. P., Mok Shim, W., Dang, S., Triantafyllou, C., & Kanwisher, N. (2008). Feedback of visual object information to foveal retinotopic cortex. *Nature Neuroscience*, 11(12), 1439–1445. <https://doi.org/10.1038/nn.2218>
- Winter, B., Perlman, M., & Majid, A. (2018). Vision dominates in perceptual language: English sensory vocabulary is optimized for usage. *Cognition*, 179, 213–220. <https://doi.org/10.1016/j.cognition.2018.05.008>
- Woolnough, O., Donos, C., Rollo, P. S., Forseth, K. J., Lakretz, Y., Crone, N. E., Fischer-Baum, S., Dehaene, S., & Tandon, N. (2020). Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nature Human Behaviour*, 1–10. <https://doi.org/10.1038/s41562-020-00982-w>
- Wu, Z., Qiu, L., Ross, A., Akyürek, E., Chen, B., Wang, B., Kim, N., Andreas, J., & Kim, Y. (2024). Reasoning or reciting? Exploring the capabilities and limitations of language models through counterfactual tasks. *arXiv*. <https://doi.org/10.48550/arXiv.2307.02477>
- Wyatte, D., Jilk, D. J., & O'Reilly, R. C. (2014). Early recurrent feedback facilitates visual object recognition under challenging conditions. *Frontiers in Psychology*, 5, 674. <https://doi.org/10.3389/fpsyg.2014.00674>
- Yamasaki, T., & Tobimatsu, S. (2018). Driving ability in alzheimer disease spectrum: Neural basis, assessment, and potential use of optic flow event-related potentials. *Frontiers in Neurology*, 9. <https://doi.org/10.3389/fneur.2018.00750>

- Yang, T., Fan, X., Hou, B., Wang, J., & Chen, X. (2024). Linguistic network in early deaf individuals: A neuroimaging meta-analysis. *NeuroImage*, 299, 120720.
<https://doi.org/10.1016/j.neuroimage.2024.120720>
- Yarkoni, T. (2020). The generalizability crisis. *Behavioral and Brain Sciences*, 45, e1.
<https://doi.org/10.1017/S0140525X20001685>
- Yeo, B. T. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., & Buckner, R. L. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3), 1125–1165. <https://doi.org/10.1152/jn.00338.2011>
- Yücel, M. A., Luke, R., Mesquita, R. C., Lühmann, A. von, Mehler, D. M. A., Lührs, M., Gemignani, J., Abdalmalak, A., Albrecht, F., Almeida, I., Artymenko, C., Ashton, K., Augustynowicz, P., Bajracharya, A., Bannier, E., Barth, B., Bayet, L., Behrendt, J., Khani, H. B., ... Zemanek, V. (2024). The fNIRS reproducibility study hub (fresh): Exploring variability and enhancing transparency in fNIRS neuroimaging research. *MetaArXiv*. <https://doi.org/10.31222/osf.io/pc6x8>
- Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: Analysis by synthesis? *Trends in Cognitive Sciences*, 10(7), 301–308. <https://doi.org/10.1016/j.tics.2006.05.002>
- Zhaoping, L. (2014). *Understanding vision: Theory, models, and data*. Oxford University Press.
<https://doi.org/10.1093/acprof:oso/9780199564668.001.0001>

Original research articles

This dissertation is based on the following original research articles:

1. **Enge, A.**, Abdel Rahman, R., & Skeide, M. A. (2021). A meta-analysis of fMRI studies of semantic cognition in children. *NeuroImage*, 241, 118436. <https://doi.org/10.1016/j.neuroimage.2021.118436>
Published under the terms of the Creative Commons CC-BY license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
2. **Enge, A.** & Skeide, M. A. (2025). Tracking the neural correlates of learning to read with dense-sampling fMRI.
Unpublished manuscript available from first author.
3. **Enge, A.**, Süß, F., & Abdel Rahman, R. (2023). Instant effects of semantic information on visual perception. *Journal of Neuroscience*, 43(26), 4896–4906. <https://doi.org/10.1523/JNEUROSCI.2038-22.2023>
Published under the terms of the Creative Commons CC-BY license.

Additional articles published during the dissertation period but not included in this thesis:

- Eiserbeck, A., **Enge, A.**, Rabovsky, M., & Abdel Rahman, R. (2021). Electrophysiological chronometry of graded consciousness during the attentional blink. *Cerebral Cortex*, bhab289. <https://doi.org/10.1093/cercor/bhab289>
- Aristei, S., Knoop, C. A., Lubrich, O., Nehrlich, T., **Enge, A.**, Stark, K., Sommer, W., & Abdel Rahman, R. (2022). Affect as Anaesthetic: How emotional contexts modulate the processing of counterintuitive concepts. *Language, Cognition and Neuroscience*. <https://doi.org/10.1080/23273798.2022.2085312>
- **Enge, A.**, Kapoor, S., Kieslinger, A.-S., & Skeide, M. A. (2023). A meta-analysis of mental rotation in the first years of life. *Developmental Science*, e13381. <https://doi.org/10.1111/desc.13381>
- Eiserbeck, A., **Enge, A.**, Rabovsky, M., & Abdel Rahman, R. (2024). Distrust before first sight? Examining knowledge- and appearance-based effects of trustworthiness on the visual consciousness of faces. *Consciousness and Cognition*, 117, 103629. <https://doi.org/10.1016/j.concog.2023.103629>
- Kessler, R., **Enge, A.**, & Skeide, M. A. (2024). How EEG preprocessing shapes decoding performance. *arXiv*. <https://doi.org/10.48550/arXiv.2410.14453>
- Knoop, C. A., Nehrlich, T., Aristei, S., Lubrich, O., Stark, K., **Enge, A.**, Sommer, W., & Abdel Rahman, R. (2024). The usual miracles: How narrative style affects the processing of counterintuitive concepts. *Psychology of Aesthetics, Creativity, and the Arts*. <https://doi.org/10.1037/aca0000730>



A meta-analysis of fMRI studies of semantic cognition in children [☆]

Alexander Enge ^{a,b,*}, Rasha Abdel Rahman ^b, Michael A. Skeide ^a



^a Research Group Learning in Early Childhood, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany
^b Department of Psychology, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489 Berlin, Germany

ARTICLE INFO

Keywords:
 fMRI
 Semantic cognition
 Children
 Meta-analysis

ABSTRACT

Our capacity to derive meaning from things that we see and words that we hear is unparalleled in other animal species and current AI systems. Despite a wealth of functional magnetic resonance imaging (fMRI) studies on where different semantic features are processed in the adult brain, the development of these systems in children is poorly understood. Here we conducted an extensive database search and identified 50 fMRI experiments investigating semantic word knowledge, semantic relatedness judgments, and the differentiation of visual semantic object categories in children (total $N = 1,018$, mean age = 10.1 years, range 4–15 years). Synthesizing the results of these experiments, we found consistent activation in the bilateral inferior frontal gyri (IFG), fusiform gyri (FG), and supplementary motor areas (SMA), as well as in the left middle and superior temporal gyri (MTG/STG). Within this system, we found little evidence for age-related changes across childhood and high overlap with the adult semantic system. In sum, the identification of these cortical areas provides the starting point for further research on the mechanisms by which the developing brain learns to make sense of its environment.

1. Introduction

The human capacity to retrieve meaning from words, phrases, and visual objects far exceeds the capacities of other animal species as well as all current state-of-the-art machine learning architectures. Functional magnetic resonance imaging (fMRI) has made it possible to map the brain areas underlying this capacity in the adult brain, showing that semantic information is processed in a distributed fashion across large parts of the cerebral cortex (Humphries et al., 2007; Huth et al., 2012, 2016; Liuzzi et al., 2020; Pulvermüller et al., 2009; Tyler et al., 2003). Most areas of this semantic system are largely amodal, that is, they show similar levels and patterns of activation regardless of whether the sensory input that is being processed comes from the visual domain or from the auditory domain (Deniz et al., 2019; Fairhall and Caramazza, 2013).

It is important, however, to interpret the findings from any individual fMRI study with caution: The generalizability of the patterns of brain activity that was observed may be limited to the respective task setting, stimuli, and population of participants under study (Yarkoni, 2021). Even when this is taken into account, the number of participants in a typical fMRI study is low (usually $N \leq 30$). Therefore, many statistically significant peaks of activation may turn out to be spurious, capitalizing on chance fluctuations in the sample rather than genuinely task-related

brain responses in the population (Button et al., 2013; Ioannidis, 2005; Thirion et al., 2007). One effective way of mitigating these two limitations is by statistically pooling the results from individual fMRI experiments on a given topic into a meta-analysis. This can be done in an image-based fashion, using the statistical parametric maps from the original experiments, or in a coordinate-based fashion, using only the peak coordinates. Despite evidence that inferences from the image-based approach are more precise (Salimi-Khorshidi et al., 2009), this approach remains difficult to implement since statistical maps for most fMRI experiments are still not being shared (Poline et al., 2012). In contrast, peak coordinates are routinely reported in research articles, oftentimes making the coordinate-based approach the only feasible one in practice (Samartsidis et al., 2017).

While many of such coordinate-based meta-analyses have been conducted for fMRI studies investigating semantic cognition in adults (Binder et al., 2009; Cocquyt et al., 2019; Ferstl et al., 2008; Jackson, 2021; Noonan et al., 2013; Rodd et al., 2015; Vigneau et al., 2006; Visser et al., 2010; Wu et al., 2012), none is available as of yet to complement this effort from a developmental perspective. One reason for this may be that children are more difficult to recruit and scan than adult participants. They often require specialized equipment, additional training sessions (e.g., familiarizing them with a mock MRI scanner),

[☆] The data and code for this study are openly available at <https://osf.io/34ry2/>. We have no conflict of interest to disclose. There were no ethical concerns since we did not collect any new data.

* Corresponding author at: Research Group Learning in Early Childhood, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany.

E-mail address: enge@cbs.mpg.de (A. Enge).

and frequently the disposal of volumes or entire runs due to excessive motion or inattentiveness. Nevertheless, a number of studies have successfully used fMRI to investigate the development of semantic processing, providing preliminary evidence for how and when the neurobiological architecture for processing meaning comes about during childhood. In these studies, children were typically scanned while probing their semantic world knowledge (e.g., by asking the child to name an object after hearing its description or to decide if a certain word refers to something animate or inanimate; e.g., Balsamo et al., 2006), their judgements of the semantic relatedness between concepts (e.g., by asking the child if two sequentially presented words or pictures were related to one another or not; e.g., Chou et al., 2019), or their viewing of different semantic categories of visual objects (e.g., by having the child perform a visual detection task while passively viewing images of human faces, tools, and scenes; e.g., Scherf et al., 2007). At this point, synthesizing these heterogeneous efforts meta-analytically is becoming important (a) to distinguish between consistent and potentially spurious findings, (b) to identify the similarities and differences between different aspects of semantic cognition (i.e., between different task categories), and (c) to identify differences in the semantic system between children and adults.

Here we conducted such a coordinate-based meta-analysis of the currently available fMRI experiments probing semantic cognition in children. Based on a systematic search of the literature using online databases, we sought to identify a wide range of fMRI studies, covering different aspects of semantic cognition and a broad age range from early childhood until the beginning of adolescence. We hypothesized that general semantic cognition would be associated with consistent activation in many of the same areas that have been found during semantic processing in adults (see, e.g., Binder et al., 2009; Jackson, 2021), namely the left inferior frontal gyrus (IFG; especially the pars triangularis and pars orbitalis), the left middle temporal gyrus (MTG) and anterior temporal lobe (ATL), as well as regions known to be sensitive to the differences between semantic visual object categories, such as the fusiform gyrus (FG) and lateral occipital complex (LOC). In children, we expected to identify additional clusters of consistent activation in the right-hemispheric homologues of these regions. This is because the left-lateralization of the language comprehension network, despite being present from newborn age onwards, continues to fully develop until early adolescence (especially in the IFG; Berl et al., 2014; Enge et al., 2020; Holland et al., 2007).

2. Materials and methods

2.1. Literature search

The search terms “(child OR children OR childhood OR pediatric) AND (brain mapping OR brain scan OR functional magnetic resonance imaging OR functional MRI OR fMRI OR neuroimaging) AND (semantics OR category OR categorization OR conceptual knowledge OR semantic knowledge OR semantic memory OR semantic feature OR semantic category OR semantic categorization OR semantic comprehension OR visual semantics OR visual categorization OR object categorization)” were entered into three online databases (PubMed/MEDLINE, PsycInfo, and Scopus). As of July 2020, this search yielded a total of 1095 articles. Of these, 895 remained after removing duplicate articles and were subsequently evaluated for eligibility (see Fig. 1). We pre-specified ten inclusion criteria, ensuring that all articles to be included (1) were written in English, (2) reported original results from a group study (excluding review articles, meta-analyses, surveys, and case studies), (3) tested at least one group of children with a mean age of 3–12 years (range 3–15 years), (4) tested a typically developing, non-clinical sample (including healthy control groups from clinical studies), (5) performed task-based fMRI (excluding resting-state fMRI and other imaging modalities), (6) had children engage in a task probing semantic cognition (i.e., semantic world knowledge, semantic relatedness, or visual object semantics),

(7) analyzed the fMRI data within the framework of the general linear model (GLM), (8) applied the same statistical threshold across the whole brain (excluding ROI analyses and partial brain coverage; Müller et al., 2018), (9) reported results as peak coordinates in standard space (Talairach or MNI), and (10) reported peaks for the within-group contrast of two semantic conditions and/or one semantic and one control condition. Initially, this led to the inclusion of 34 articles. We consulted the introduction and reference sections of these articles as well as relevant review papers on children's semantic and language processing (Antonucci and Alt, 2011; Barqueró et al., 2014; Enge et al., 2020; Leach and Holland, 2010; Martin et al., 2015; O'Shaughnessy et al., 2008; Sachs and Gaillard, 2003; Schlaggar and McCandliss, 2007; Skeide and Friederici, 2016; Weiss-Croft and Baldeweg, 2015) to identify additional articles not covered by our database search. Following this procedure, we identified 23 additional articles, seven of which fulfilled all inclusion criteria.

In addition to those articles fulfilling all ten inclusion criteria, 37 articles met all but the last criteria—that is, the relevant within-group peak coordinates were not reported in the published article. In these cases, we contacted the corresponding authors to request the missing information. This led to the inclusion of three additional articles, resulting in a total of 45 articles being included in the meta-analysis.

Whenever one of these articles reported multiple contrasts based on the same sample of children, the coordinates from all of these contrasts were treated as a single experiment (note that *experiment* is the term we use whenever we refer to our primary unit of analysis; Turkeltaub et al., 2012). This is considered good practice in order to minimize within-group effects and avoid inflating the number of independent data points included in the meta-analysis (Eickhoff et al., 2012; Turkeltaub et al., 2012). Conversely, whenever an article reported multiple contrasts from two or more independent samples of children, these were treated as separate experiments. This led to a final meta-analytic sample of $n = 50$ experiments. While each of these experiments targeted some aspect of general semantic cognition, they could further be subdivided into more homogeneous groups of experimental tasks, probing (a) semantic world knowledge (e.g., naming an object after hearing its description; $n = 21$), (b) semantic relatedness (e.g., hearing two words and deciding if they are related or not; $n = 16$), and (c) visual semantic object categories (e.g., passively viewing faces as compared to other visual stimuli; $n = 13$). The experiments belonging to these three task categories as well as the entire data set (probing general semantic cognition across all task categories) were meta-analyzed using two different algorithms: activation likelihood estimation and seed-based *d* mapping (see Fig. 2).

2.2. Activation likelihood estimation

Activation likelihood estimation (ALE) is the most frequently used algorithm to perform coordinate-based meta-analyses of neuroimaging experiments (Acar et al., 2018). It estimates the degree to which peak coordinates taken from independent MRI experiments, all investigating the same task and/or participant population, spatially converge to form non-random clusters of activation (Eickhoff et al., 2009, 2012; Turkeltaub et al., 2002). To this end, the algorithm first recreates a modeled activation map for each of the input experiments. All voxels for which the experiment reports a peak coordinate are assigned a value of 1, whereas all of the other voxels within the gray matter mask are assigned a value of 0. Because these peaks entail spatial uncertainty and are assumed to be part of larger clusters of activation, their values are smoothed across the neighboring voxels by convolving them with a Gaussian kernel. The width of this kernel is chosen to be inversely proportional to the sample size of the experiment, reflecting the fact that larger sample sizes provide stronger evidence for the true location of any peak of activation (Eickhoff et al., 2009). When two (or more) peaks in the experiment are reported in close proximity to one another, the voxels at which their Gaussians overlap are assigned the max-

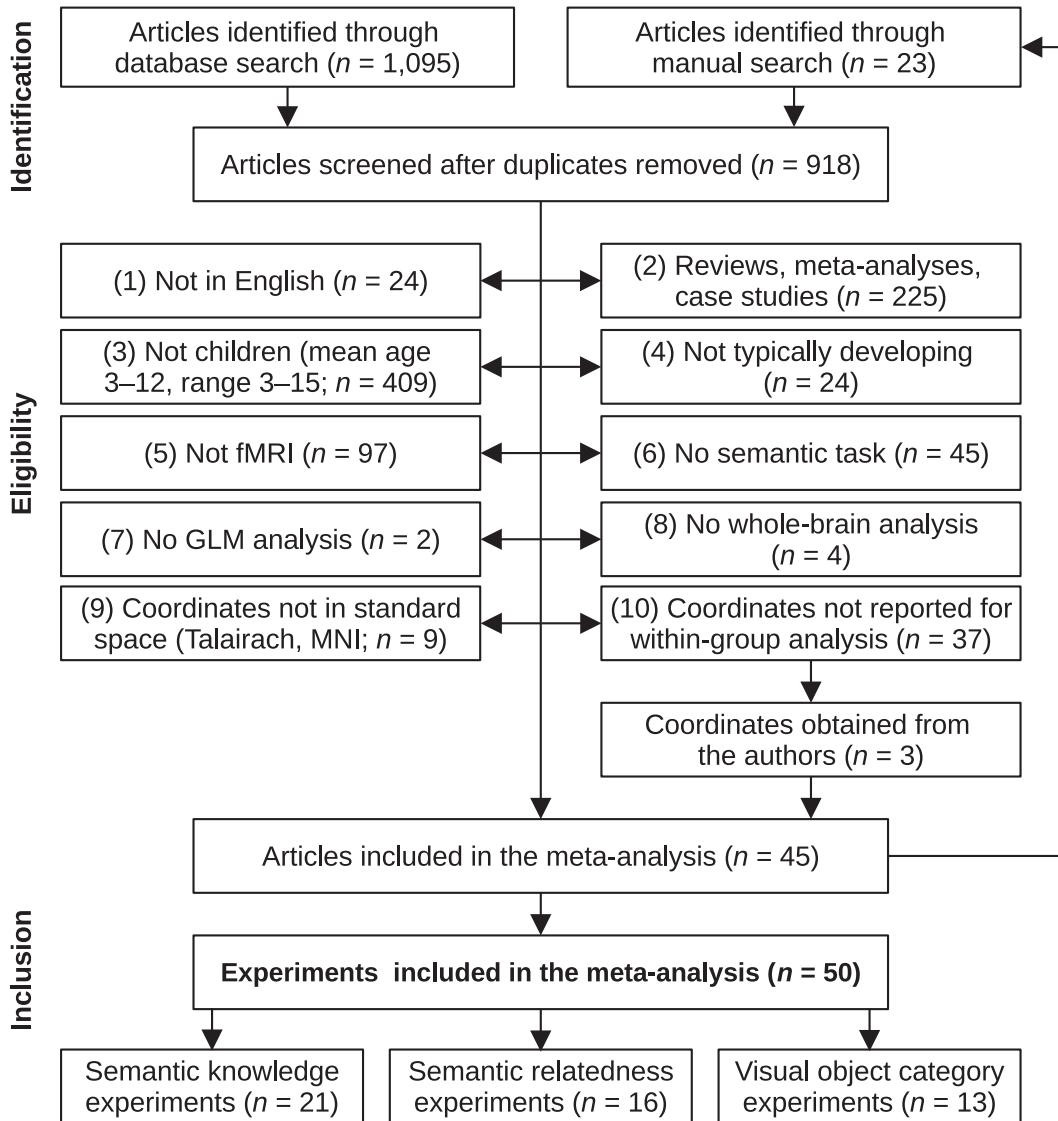


Fig. 1. Literature search and selection workflow.

imum—rather than the sum—of their respective values. This prevents any meta-analytic cluster from receiving artificially high likelihood values merely because a large number of subpeaks has been reported for this cluster in the original article.

The modeled activation maps of the individual experiments are then combined into a single meta-analytic map by assigning an ALE value to each voxel. This ALE value is computed as the union of the modeled activation values for this voxel across the modeled activation maps for all k experiments (Acar et al., 2018; Eickhoff et al., 2012):

$$ALE_{xyz} = 1 - \prod_{i=1}^k (1 - MA_{kxyz}).$$

This hierarchical procedure treats the included experiments as a random subsample of all possible experiments and therefore allows the gen-

eralization across the population of possible fMRI experiments on the topic of interest (Eickhoff et al., 2009). The statistical significance of these voxel-wise ALE values is determined by comparing them to an analytically derived null distribution as described by Eickhoff et al. (2012). To correct for multiple comparisons, a cluster-level family-wise error (FWE) correction procedure has been shown to offer an excellent trade-off between control over the Type I error rate and statistical power (Eickhoff et al., 2016).

For the present analysis, these steps were performed using the Ni-MARE package (Version 0.0.9; Salo et al., 2020) in Python (Version 3.8.8; Van Rossum and Drake, 2009). If necessary, coordinates were transformed from Talairach to MNI space using the icbm2tal transform function (Lancaster et al., 2007). The modeled activation maps were rendered in MNI152 space at $2 \times 2 \times 2$ mm resolution (Fonov et al., 2011).

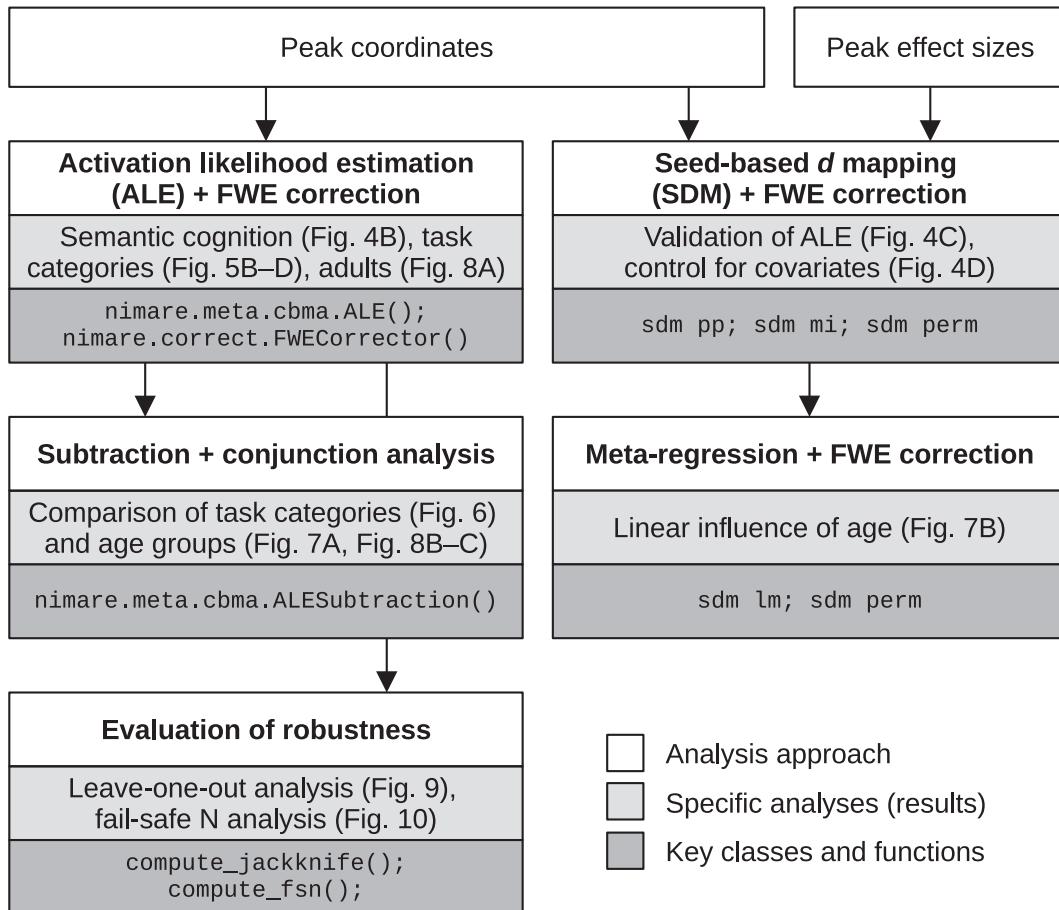


Fig. 2. Meta-analytic approaches.

Two meta-analytic approaches (activation likelihood estimation and seed-based d mapping) were used to determine the overlap of peak fMRI activations associated with semantic cognition in children. Subtraction analysis, conjunction analysis, and meta-regression were used to test the influence of moderating variables (e.g., semantic task category, mean age) and to compare our results to previous findings obtained in adult samples. Finally, leave-one-out and fail-safe N analyses were carried out to probe the robustness of the meta-analytic results against different types of publication bias. The Python code for all analyses is available in reproducible notebooks under <https://osf.io/34ry2/>. FWE = family-wise error, Fig. = Figure.

For statistical thresholding, a voxel-level cluster-forming threshold of $p < .001$ (uncorrected) and a cluster-level threshold of $p < .01$ (FWE-corrected) was used. This cluster-level threshold was determined by comparing the observed cluster size to an empirical distribution built from 1000 iterations of drawing random peak locations from the gray matter template and recording the maximal cluster size. The Nilearn package (Version 0.7.1; [Abraham et al., 2014](#)) was used for image processing and plotting while the anatomic automatic labeling atlas (AAL2; [Rolls et al., 2015](#)) as implemented in the AtlasReader package (Version 0.1.2; [Notter et al., 2019](#)) was used for anatomical labeling.

2.3. Seed-Based d mapping

While ALE estimates the spatial convergence of reported activation peaks, an alternative approach is to use the effect sizes (if available) of these peaks to infer a meta-analytic effect size for each gray matter voxel. This approach more closely resembles traditional meta-analyses of behavioral or clinical outcomes and is used by the seed-based d map-

ping (SDM) algorithm ([Albajes-Eizagirre et al., 2019](#)). In short, it determines lower and upper bounds for possible effect size images based on the peak coordinates and their reported effect sizes (t scores or z scores). Then, a meta-analytic method correcting for non-statistically significant unreported effects (MetaNSUE; [Albajes-Eizagirre et al., 2019](#)) is used to infer the most plausible effect size and its standard error based on multiple imputations of censored information. Subsequently, all imputed data sets are meta-analyzed separately and combined using Rubin's rules. For statistical thresholding, the resulting meta-analytic map is FWE-corrected by comparing the voxel-wise observed effect size against an empirical null distribution of effect sizes built from random permutations.

The SDM algorithm was used on the same experiments and peak coordinates as for the ALE analysis but adding, if available, their reported t scores or z scores (the latter being converted to t scores with $df = n_{\text{children}} - 1$). The data were preprocessed with the SDM-PSI software (Version 6.21, <https://www.sdmproject.com>), using its default gray matter correlation template with a voxel size of $2 \times 2 \times 2$ mm and

a Gaussian smoothing kernel (anisotropy $\alpha = 1.0$, FWHM = 20 mm). The effect-size based SDM algorithm made it possible (a) to probe the robustness of the ALE results against a change in the meta-analytic approach and (b) to statistically control for systematic differences between the included experiments by means of a covariate analysis. Note that the latter type of analysis is impossible in an approach like ALE because it disregards the effect sizes of the reported peaks and instead treats them as binary. Two separate models were computed to meet these two objectives: (a) a mean-based meta-analysis without any covariates or predictors (as in the ALE analysis) and (b) a mean-based meta-analysis controlling for four different experiment-level confounds (language of the experiment [0 = German, 1 = Dutch, 2 = French, 3 = English, 4 = Japanese, 5 = Mandarin Chinese], presentation modality [0 = visual, 1 = audiovisual, 2 = auditory and visual in separate blocks, 3 = auditory], response modality [0 = no response, 1 = manual response, 2 = covert speech, 3 = overt speech], and data analysis software [0 = SPM, 1 = FSL, 2 = other]). Both models were estimated with 50 random imputations and statistical thresholding was performed using a voxel-level FWE-corrected threshold of $p < .001$ and a cluster extent threshold of $k > 25$ connected voxels (200 mm³).

2.4. Differences between semantic task categories

Beyond mapping the cortical network associated with general and task-specific semantic cognition in children, a meta-analytic subtraction analysis (Laird et al., 2005) was carried out to test for reliable differences between the three different task categories (i.e., knowledge, relatedness, and objects). For this type of analysis, one ALE map (e.g., the map for the semantic knowledge experiments) was subtracted from another ALE map (e.g., the combined map for the semantic relatedness and visual object category experiments). The resulting map of difference scores was then compared against an empirical null distribution of such difference maps, obtained from randomly reshuffling the original experiments 20,000 times into new groups. Voxels with $p < .001$ (uncorrected) as compared to this null distribution and forming clusters of at least 25 connected voxels (200 mm³) were considered as showing reliable differences between task categories. In addition, a meta-analytic conjunction analysis was performed to identify areas where cognitive processing was shared across all three task categories. This was done by taking the minimum ALE value at each voxel across all three task-specific ALE maps—but only for those voxels that were statistically significant in each of the three (Nichols et al., 2005).

2.5. Age-related changes

The same approach as just described was used to compare semantic cognition in older versus younger children. To this end, the original sample was split into equally sized groups at the median of the (mean) sample ages across experiments. These two groups of experiments were compared using the same subtraction procedure and statistical threshold as for the semantic task categories. Additionally, we also tested for a linear influence of age by means of a meta-regression using SDM (see Section 2.3). In this linear model, the outcome of interest was not the voxel-wise effect size across experiments (as in the main SDM analysis) but those voxels whose effect size showed significant covariation with the (mean) age of the sample(s) of children contributing to it.

2.6. Comparison with semantic cognition in adults

The meta-analytic results of semantic cognition in children were also compared to semantic cognition in adults as reported in a recent meta-analysis on semantic control (Jackson, 2021). To this end, we recreated their ALE analysis of general semantic cognition in adults ($n = 415$ experiments; see their Fig. 3 and Table 3) and compared it to our child-specific ALE analysis by means of a meta-analytic subtraction and conjunction analysis as described above (see Section 2.4).

2.7. Evaluation of robustness

Meta-analyses reflect the state of the published literature on a given topic and are therefore subject to the same biases as the original studies (e.g., small sample bias, selective reporting, file drawer problem). For behavioral and clinical meta-analyses, a range of standard tools has been developed to assess the risk of these biases as well as the robustness of the meta-analytic results against them. Some but not all of these tools can be carried over to the meta-analysis of neuroimaging data (Acar et al., 2018). For instance, it is possible to assess the degree to which the meta-analytic results depend on any individual study, which may or may not have reported false positive findings (e.g., due to low statistical power; Button et al., 2013; Ioannidis, 2005; Thirion et al., 2007). This can be done by recomputing the original meta-analysis as many times as there are experiments included, each time leaving out one of these experiments. This leave-one-out analysis (also called jackknife analysis) reveals if any of the observed meta-analytic clusters critically depends on a single influential experiment or if it is robust against a false positive experiment in the sample.

Publication bias may not only manifest itself in the form of published experiments reporting false positive effects but also in the form of experiments not getting published when failing to obtain statistically significant effects. Because of this “file drawer” problem, there are up to approximately 30 unpublished neuroimaging experiments with null effects (i.e., reporting zero significant peaks) per 100 published experiments (Samartsidis et al., 2020). While these cannot be factored into the meta-analysis directly, the simulation of imaginary file drawers with different numbers of null experiments is informative regarding the robustness of the results against this type of bias. In this context, the fail-safe N (FSN) metric has been defined as the number of null experiments that can be added to the original meta-analysis without rendering its meta-analytic effect size statistically non-significant (Rosenthal, 1979). If FSN exceeds the upper bound of experiments that can realistically be expected to be inside the file drawer, one can conclude that the file drawer problem does not suffice to explain the meta-analytic result. This logic can be extended to meta-analyses of neuroimaging studies by simulating null experiments with peaks of activation at random rather than spatially converging locations across the brain (Acar et al., 2018). Such null experiments were simulated as to resemble the original experiments in terms of their individual sample sizes and numbers of reported peak coordinates but had their peak locations drawn randomly from all possible voxels within the gray matter template. They were then added iteratively to the original experiments. At each step, the ALE analysis was repeated, recording for every voxel if it had remained part of a statistically significant cluster or not. This was repeated up to a maximum of five times the number of experiments in the original sample (e.g., FSN_{max} = 150 for our main analysis with $n = 50$ experiments). The whole procedure was repeated for 10 different (random) file drawers of null experiments.

Both of these approaches (leave-one-out and fail-safe N analysis) were performed separately for our main ALE analysis (including all 50 experiments) as well as for the task category-specific sub-analyses. They were expected to be especially informative for the category-specific analyses because the low number of experiments in these sub-analyses might have reduced the robustness of the meta-analytic results (Eickhoff et al., 2016).

3. Results

3.1. Literature search

As of July 2020, a total of 45 articles reporting 50 independent fMRI experiments of semantic cognition in children were obtained by searching online literature databases (see Fig. 1 and Table 1). These experiments could be grouped further into experiments probing semantic world knowledge ($n = 21$), semantic relatedness judgments ($n = 16$),

Table 1
Experiments included in the meta-analysis.

#	Author(s) (year)	n	Sex (F/M)	Hand. (R/L)	Mean age (SD)	Age range	Semantic task	Task category	No. of peaks
1	Arnoldussen (2006)	11	NA	NA	11.6 (1.0)	9–13	Picture–word matching vs. shape matching	Knowledge	12
2	Arnoldussen (2006)	11	NA	NA	7.9 (0.6)	7–8	Picture–word matching vs. shape matching	Knowledge	11
3	Aylward et al. (2005)	11	4/7	10/1	9.3 (0.9)	8–10	Viewing faces vs. viewing houses	Objects	6
4	Backes et al. (2002)	8	0/8	8/0	11.6 (0.7)	11–12	Animal/no-animal judgment vs. fixation	Knowledge	8
5	Balsamo et al. (2002)	11	7/4	11/0	8.5 (0.9)	7–9	Naming after description vs. rest	Knowledge	26
6	Balsamo et al. (2006)	23	13/10	22/1	8.5 (1.5)	5–10	Category–word judgment vs. tone detection	Knowledge	9
7	Bauer et al. (2017)	14	7/7	14/0	10.3 (0.9)	8–11	Object size or animacy judgment vs. fixation	Knowledge	12
8	Berl et al. (2014)	57	26/31	57/0	8.9 (NA)	4–12	Definition–word matching vs. tone detection	Knowledge	15
9	Booth et al. (2001)	5	0/5	5/0	11.1 (NA)	9–12	Word relatedness judgement vs. tone/symbol matching	Relatedness	9
10	Booth et al. (2003)	15	8/7	15/0	10.7 (NA)	9–11	Word relatedness judgement vs. tone/symbol matching	Relatedness	4
11	Booth et al. (2007)	13	4/9	13/0	10.5 (2.2)	9–15	Word relatedness judgement vs. false font matching	Relatedness	15
12	Brauer & Friederici (2007)	12	8/4	12/0	6.2 (NA)	5–6	Semantic sentence acceptability judgment vs. rest	Knowledge	24
13	Cao et al. (2008)	13	6/7	13/0	11.2 (NA)	9–12	Word relatedness judgement vs. symbol string matching	Relatedness	9
14	Chou et al. (2006a)	35	22/13	35/0	11.7 (2.1)	9–15	Word relatedness judgment vs. color change detection	Relatedness	20
15	Chou et al. (2006b)	26	NA	26/0	12.1 (2.0)	9–15	Word relatedness judgment vs. color change detection	Relatedness	18
16	Chou et al. (2009)	33	16/17	33/0	12.3 (1.8)	8–15	Word relatedness judgment vs. non-character matching	Relatedness	20
17	Chou et al. (2019)	16	5/11	16/0	12.1 (1.4)	10–14	Word relatedness judgement (multiple contrasts)	Relatedness	5
18	Corbett et al. (2009)	15	2/13	NA	9.2 (1.4)	8–12	Sequential matching of faces vs. things	Objects	4
19	Dekker et al. (2014)	10	2/8	10/0	9.8 (0.4)	9–10	Categorical one-back task (multiple categories)	Objects	10
20	Dekker et al. (2014)	11	4/7	11/0	7.6 (0.4)	7–8	Categorical one-back task (multiple categories)	Objects	7
21	Fan et al. (2020)	26	14/12	26/0	9.8 (1.5)	8–12	Word relatedness judgment vs. visual cue detection	Relatedness	4
22	Gaillard et al. (2001)	9	4/5	9/0	10.2 (NA)	7–13	Naming after description vs. dot pattern viewing	Knowledge	4
23	Gaillard et al. (2003)	16	9/7	16/0	10.2 (NA)	7–14	Word generation to category names vs. rest	Knowledge	13
24	Horowitz-Kraus et al. (2015)	23	15/8	23/0	8.5 (0.8)	7–9	Sentence–picture matching vs. word–picture matching	Knowledge	28
25	Kersey et al. (2016)	29	12/17	NA	6.6 (NA)	4–8	Picture matching with tools vs. other categories	Objects	30
26	Krishnan et al. (2015)	37	19/18	37/0	9.7 (NA)	7–12	Picture matching vs. saying “silly”	Knowledge	5
27	Krishnan et al. (2021)	67	NA	NA	12.1 (1.7)	10–15	Verb generation to pictures vs. rest	Knowledge	22
28	Lee et al. (2011)	23	11/12	23/0	12.8 (1.5)	10–15	Word relatedness judgment vs. word–tone matching	Relatedness	21
29	Lee et al. (2011)	23	11/12	23/0	12.8 (1.5)	10–15	Word relatedness judgment vs. non-character matching	Relatedness	16
30	Lee et al. (2016)	30	14/16	30/0	11.8 (1.9)	8–15	Word relatedness judgment vs. non-character matching	Relatedness	6
31	Libertus et al. (2009)	15	7/8	NA	8.7 (NA)	8–9	Two-back task with faces vs. other categories	Objects	12
32	Liebig et al. (2017)	41	18/23	41/0	11.9 (NA)	9–13	Animacy judgment vs. visual symbol judgment	Objects	15
33	Mathur et al. (2020)	19	11/8	19/0	6.6 (NA)	5–7	Word–picture relatedness judgment vs. symbol string matching	Relatedness	10
34	Meyler et al. (2008)	12	9/3	12/0	10.8 (0.4)	NA	Semantic sentence acceptability judgment vs. fixation	Knowledge	13
35	Monzalvo et al. (2012)	23	11/12	NA	9.6 (0.5)	8–10	Target detection while viewing faces/houses vs. other categories	Objects	10
36	Moore-Parks et al. (2010)	23	12/11	23/0	8.8 (1.1)	7–10	Definition–word matching vs. reversed speech	Knowledge	15
37	Okamoto et al. (2017)	12	1/11	11/1	11.3 (1.3)	NA	Viewing bodies/faces vs. other categories	Objects	10
38	Passarotti et al. (2003)	12	8/4	12/0	NA (NA)	10–12	Face matching vs. detecting scrambled faces	Objects	15
39	Schafer et al. (2009)	26	15/11	24/2	12.2 (0.4)	NA	Word relatedness judgment vs. visual symbol judgment	Relatedness	8
40	Scherf et al. (2007)	10	4/6	10/0	12.5 (1.0)	11–14	Viewing movies of faces/places/objects vs. other categories	Objects	34
41	Scherf et al. (2007)	10	4/6	10/0	7.2 (1.0)	5–8	Viewing movies of faces/places/objects vs. other categories	Objects	24
42	Scherf et al. (2010)	10	0/10	10/0	12.4 (1.3)	10–14	Viewing movies of faces/places/objects vs. other categories	Objects	47
43	Siok et al. (2004)	8	4/4	8/0	11.1 (NA)	10–12	Chinese character judgment vs. fixation	Knowledge	17
44	Skeide et al. (2014)	20	8/12	20/0	10.3 (NA)	9–10	Picture–sentence matching (plausible vs. implausible)	Knowledge	1
45	Skeide et al. (2014)	20	11/9	20/0	7.4 (NA)	6–7	Picture–sentence matching (plausible vs. implausible)	Knowledge	1
46	Szaflarski et al. (2006)	29	15/14	NA	NA (NA)	5–11	Verb generation to nouns vs. finger tapping to tones	Knowledge	10
47	Vannest et al. (2012)	15	6/9	15/0	9.2 (NA)	7–14	Animal judgement vs. tone detection	Knowledge	4
48	Wong et al. (2019)	38	0/38	38/0	11.9 (1.0)	NA	Word relatedness judgement (multiple contrasts)	Relatedness	18
49	Wu et al. (2016)	30	20/10	30/0	5.5 (0.3)	5–5	Sentence listening (prototypical, neutral, non-prototypical)	Knowledge	12
50	Xue et al. (2004)	12	6/6	12/0	11.6 (NA)	10–12	Word relatedness judgment vs. fixation	Relatedness	18

Note. Multiple experiments were derived from the same original article if and only if they were testing independent groups of children (see Section 2.1 and Turkeltaub et al., 2012). Additional information about the experiments can be found in Appendix A. # = experiment ID, n = sample size, F = female, M = male, hand. = handedness, R = right-handed, L = left-handed, no. = number, NA = not available.

Table 2
Statistics of the meta-analytic clusters shown in Fig. 4.

Analysis	#	Size (mm ³)	Mean z	Mean ALE	Peak z	Peak ALE	Peak X	Peak Y	Peak Z	Peak anatomical label
Activation like-hood estimation	1	10,232	3.93	0.022	6.45	0.043	-44	18	24	L inferior frontal gyrus (tri.)
	2	5736	4.44	0.026	7.43	0.053	-4	16	50	L supplementary motor area
	3	3312	3.67	0.020	4.84	0.029	-40	-52	-20	L fusiform gyrus
	4	2872	4.46	0.027	7.49	0.053	36	22	-6	R insula
	5	2504	4.37	0.026	7.05	0.049	-52	-38	4	L middle temporal gyrus
	6	1856	3.52	0.019	4.24	0.024	36	-90	-6	R inferior occipital gyrus
	7	1464	3.64	0.020	4.88	0.029	40	-52	-18	R fusiform gyrus
	8	88	3.23	0.017	3.39	0.018	-38	36	6	L inferior frontal gyrus (tri.)
Seed-based d mapping	1	22,672	5.81		7.56		-4	-4	56	L supplementary motor area
	2	20,568	6.13		9.62		-46	16	-10	L inferior frontal gyrus (orb.)
	3	4696	5.50		7.13		-54	-52	8	L middle temporal gyrus
	4	3912	5.40		6.43		48	20	4	R inferior frontal gyrus (tri.)
	5	2632	5.36		6.19		-32	-52	-18	L fusiform gyrus
	6	2568	5.42		6.59		28	-46	-26	R cerebellum
	7	576	5.43		6.08		-40	-66	42	L angular gyrus
	8	552	5.37		6.21		-44	4	48	L precentral gyrus
	9	416	5.24		5.73		-18	-42	-16	L fusiform gyrus
Seed-based d mapping + covariates	1	71,648	6.04		9.76		-48	24	-8	L inferior frontal gyrus (orb.)
	2	30,912	6.22		8.15		0	0	48	L supplementary motor area
	3	21,712	5.77		7.30		36	12	2	R insula
	4	16,696	5.74		7.41		46	-56	-12	Right inferior temporal gyrus
	5	3464	5.59		6.66		12	-90	6	Right calcarine sulcus
	6	2928	5.60		6.46		-26	-78	36	Left middle occipital gyrus
	7	2160	5.53		6.07		-4	-48	32	Left posterior cingulate cortex

Note. Peak anatomical labels are based on the anatomic automatic labeling atlas (AAL2; [Rolls et al., 2015](#)). # = cluster ID, L = left, R = right, tri. = pars triangularis, orb. = pars orbitalis.

and the discrimination of visual semantic object categories ($n = 13$). Together, they comprise fMRI data of 1018 children ($m = 20.4$ per experiment, $md = 15.5$, range 5–67; see Fig. 3) with a mean age of 10.1 years (range of mean ages 5.5–12.8 years, total age range 4–15 years). According to the original articles, 54.4% of these children were boys (45.6% were girls) and 98.4% were right-handed (1.6% were left-handed). From these experiments, a total of 687 peaks of activation were reported ($m = 13.7$ per experiment, $md = 12$, range 1–47) and entered into the meta-analysis. Of these peaks, 400 (58.2%) were in the left hemisphere ($x_{MNI} < 0$), indicating a slight degree of lateralization. There were no reliable associations across experiments between sample size, the mean age of the children under study, and the number of peaks reported (see Fig. 3). Additional descriptive information about the experiments can be found in Appendix A.

3.2. Activation likelihood estimation

For general semantic cognition in children, the meta-analysis using ALE revealed spatial convergence of activation across experiments in eight clusters distributed across different regions of children's cortex (see Table 2 and Fig. 4B). Ordered by cluster size, they were located in the left inferior frontal and precentral gyri (Clusters #1 and #8), the bilateral supplementary motor areas (Cluster #2), the left fusiform gyrus (Cluster #3), the right insular and inferior frontal cortices (Cluster #4), the left middle and superior temporal cortices (Cluster #5), the right inferior occipital gyrus and calcarine sulcus (Cluster #6), and the right fusiform gyrus (Cluster #7).

3.3. Seed-based d mapping

Meta-analytic effect size maps were created based on the test statistics (z scores or t scores) of the reported peak coordinates. These were available for 461 (67.1%) of all peak coordinates, whereas the test statistics for the remaining 226 peak coordinates were inferred via multiple imputations as described in [Albajes-Eizagirre et al. \(2019\)](#). This alternative meta-analytic approach yielded qualitatively similar results as ALE: The largest clusters (and highest effect sizes) were again observed in the left inferior frontal gyrus, the bilateral supplementary motor areas, and

the left middle and superior temporal gyri (see Fig. 4C and Table 2). Three noteworthy differences between the results from these two different meta-analytic algorithms were (a) that the size of the significant clusters was larger overall for SDM as compared to ALE, (b) that one cluster in the right visual cortex was observed with ALE but not with SDM, and (c) that one cluster in the left angular gyrus was observed with SDM but not with ALE.

The effect size-based approach not only served as a robustness check for the main (ALE) analysis but also made it possible to re-assess the results while controlling for four different linear covariates of no interest (namely the language of the experiment, the modality of stimulus presentation, the modality of children's response, and the statistical software package used for data analysis). This again yielded qualitatively similar results, although cluster sizes were larger than in the original analysis without covariates (see Fig. 4D and Table 2).

3.4. Differences between semantic task categories

Task category-specific sub-analyses for experiments probing semantic world knowledge (e.g., naming a word after hearing its description) and for experiments probing semantic relatedness (e.g., hearing two words and deciding if they are related or not) both showed the largest clusters of activation in the bilateral supplementary motor areas, in the pars triangularis of the left inferior frontal gyrus, and in the right insular and inferior frontal cortices (see Fig. 5 and Table 3). For experiments probing semantic relatedness, there was one additional cluster in the left middle temporal gyrus. The sub-analysis for experiments probing the discrimination of visual semantic object categories showed three clusters of consistent activation in the bilateral fusiform gyri as well as in the visual cortex of the right occipital lobe.

These task-specific meta-analytic maps were contrasted against one another to examine where they differed reliably from one another (see Fig. 6 and Table 4). First, experiments probing semantic knowledge showed more consistent activation than the other two task categories in two small clusters in the left insular and middle frontal cortices. Second, tasks probing semantic relatedness showed more consistent activation than the other two task categories in the pars opercularis of the left inferior frontal gyrus and in the left middle temporal gyrus. Finally,

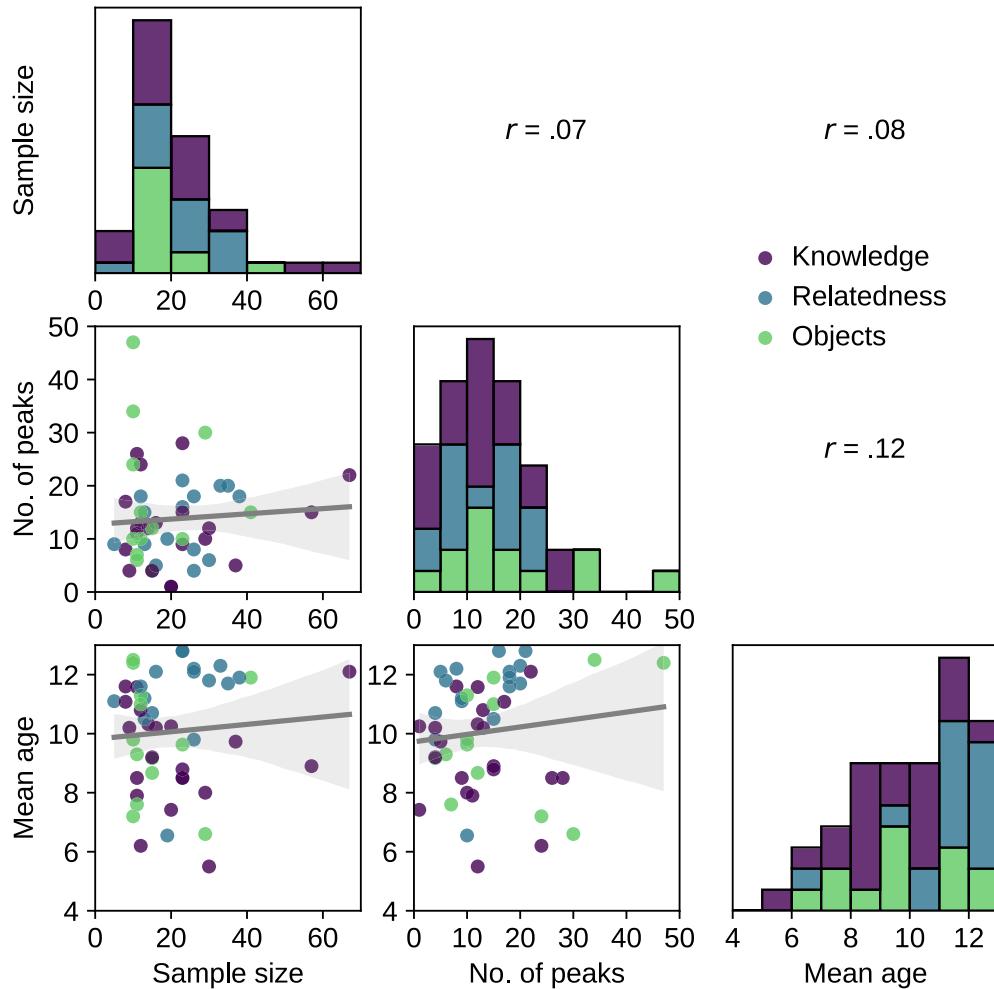


Fig. 3. Distributions and bivariate associations of experiment-level characteristics.

Histograms on the main diagonal show the number of experiments in the meta-analysis, binned according to their sample size, number of peaks reported, and mean age of the children under study. Scatterplots and correlation coefficients on the off-diagonals show the bivariate associations between these experiment-level characteristics. The gray lines show the linear regression trends with their 95% confidence interval. No. = number.

Table 3
Statistics for the meta-analytic clusters shown in Fig. 5.

Analysis	#	Size (mm ³)	Mean z	Mean ALE	Peak z	Peak ALE	Peak X	Peak Y	Peak Z	Peak anatomical label
Knowledge	1	3472	4.06	0.017	6.76	0.036	-4	16	50	L supplementary motor area
	2	3096	3.88	0.016	5.52	0.027	-44	16	24	L inferior frontal gyrus (tri.)
	3	1128	3.69	0.015	4.81	0.022	36	24	-8	R inferior frontal gyrus (orb.)
	4	128	3.27	0.012	3.61	0.014	-56	20	22	L inferior frontal gyrus (tri.)
Relatedness	1	6048	4.00	0.016	6.45	0.032	-48	22	10	L inferior frontal gyrus (tri.)
	2	3480	4.01	0.016	6.61	0.034	-2	8	60	L supplementary motor area
	3	2760	4.13	0.017	6.60	0.034	-54	-42	4	L middle temporal gyrus
	4	1888	3.97	0.016	5.74	0.027	36	22	-6	R insula
Objects	1	3408	3.66	0.013	4.99	0.020	42	-52	-20	R fusiform gyrus
	2	2720	3.67	0.013	4.71	0.019	-40	-52	-20	L fusiform gyrus
	3	1800	3.62	0.013	4.95	0.020	24	-92	-4	R inferior occipital gyrus

Note. Peak anatomical labels are based on the anatomic automatic labeling atlas (AAL2; Rolls et al., 2015). # = cluster ID, L = left, R = right, tri. = pars triangularis, orb. = pars orbitalis.

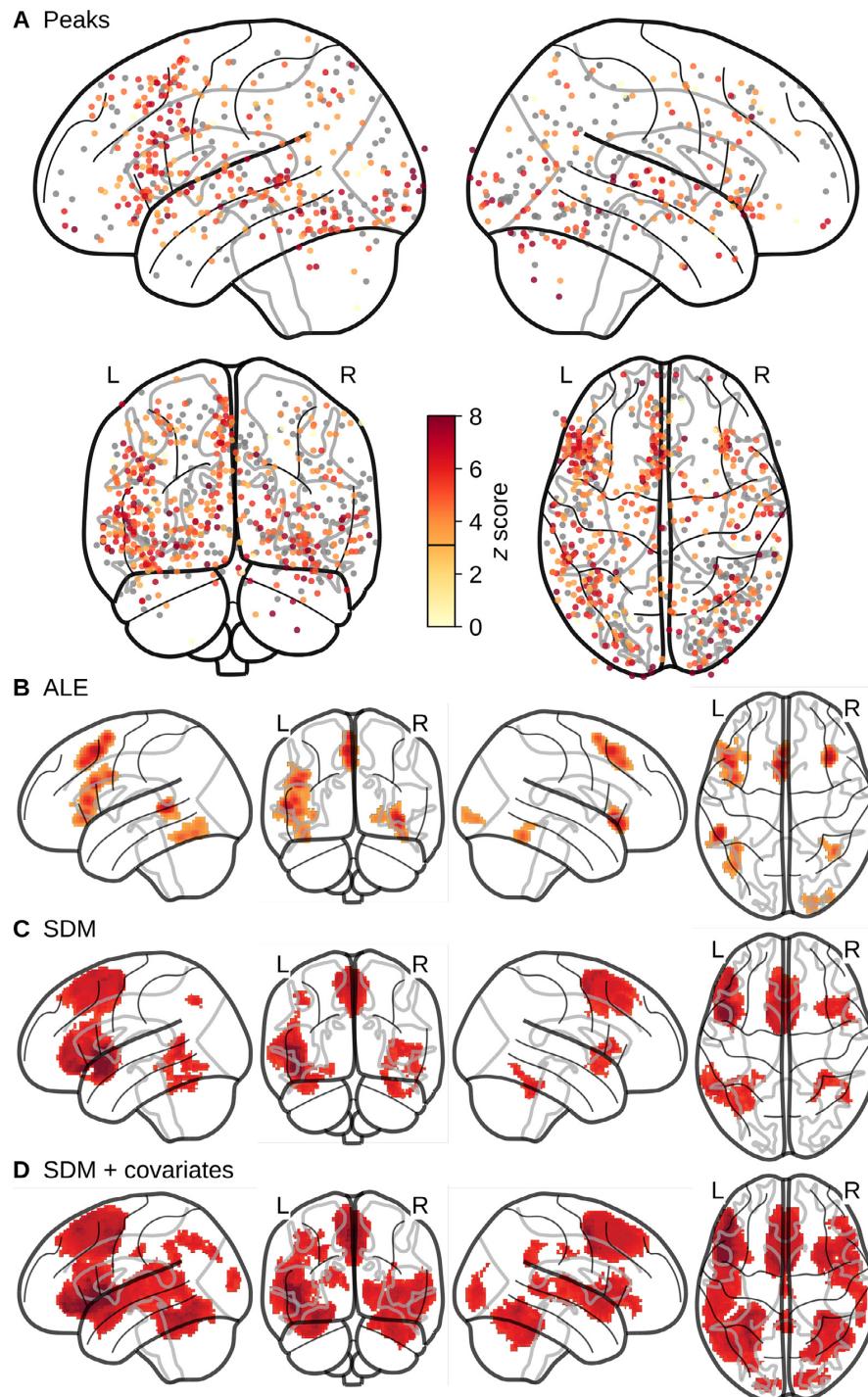


Fig. 4. Meta-analytic results for semantic cognition in children.

(A) A total of 687 individual peaks from 50 fMRI experiments of semantic cognition in children are shown together with their test statistic (converted to z scores) as reported in the original articles. Peaks for which no test statistic was reported are shown in gray. (B) Meta-analytic clusters with above-chance overlap revealed by activation likelihood estimation (ALE), thresholded at $p < .001$ (uncorrected) at the voxel level and $p < .01$ (FWE-corrected) at the cluster level. (C) Meta-analytic clusters with above-chance effect sizes from seed-based d mapping (SDM), thresholded at $p < .001$ (FWE-corrected) at the voxel level and $k > 25$ connected voxels (200 mm^3) at the cluster level. (D) The same SDM analysis and statistical thresholding but controlling for four linear covariates of no interest, namely the language of the experiment, the modality of stimulus presentation, the modality of children's responses, and the statistical software package used in the original article. See Appendix B for an exploratory assessment of the unique influence of each of these covariates using ALE subtraction analyses (as described in Section 2.4).

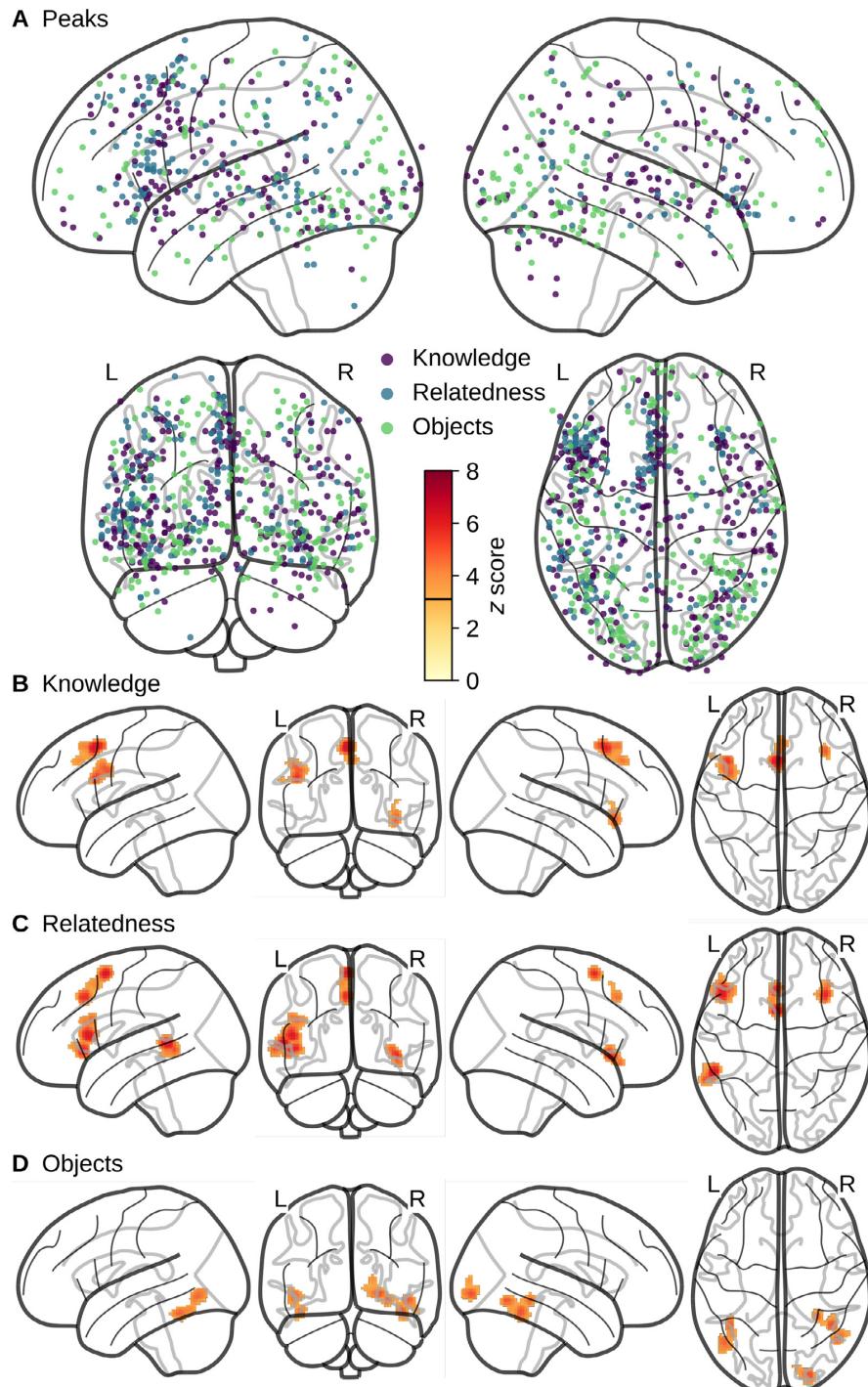


Fig. 5. Sub-analyses for three semantic task categories.

The individual peaks in (A) are shown with color-coding representing the category of semantic task (purple: knowledge, blue: relatedness, green: objects). The clusters derived from activation likelihood estimation for (B) semantic knowledge experiments, (C) semantic relatedness experiments, and (D) visual semantic object category experiments are each thresholded at $p < .001$ (uncorrected) at the voxel level and $p < .01$ (FWE-corrected) at the cluster level.

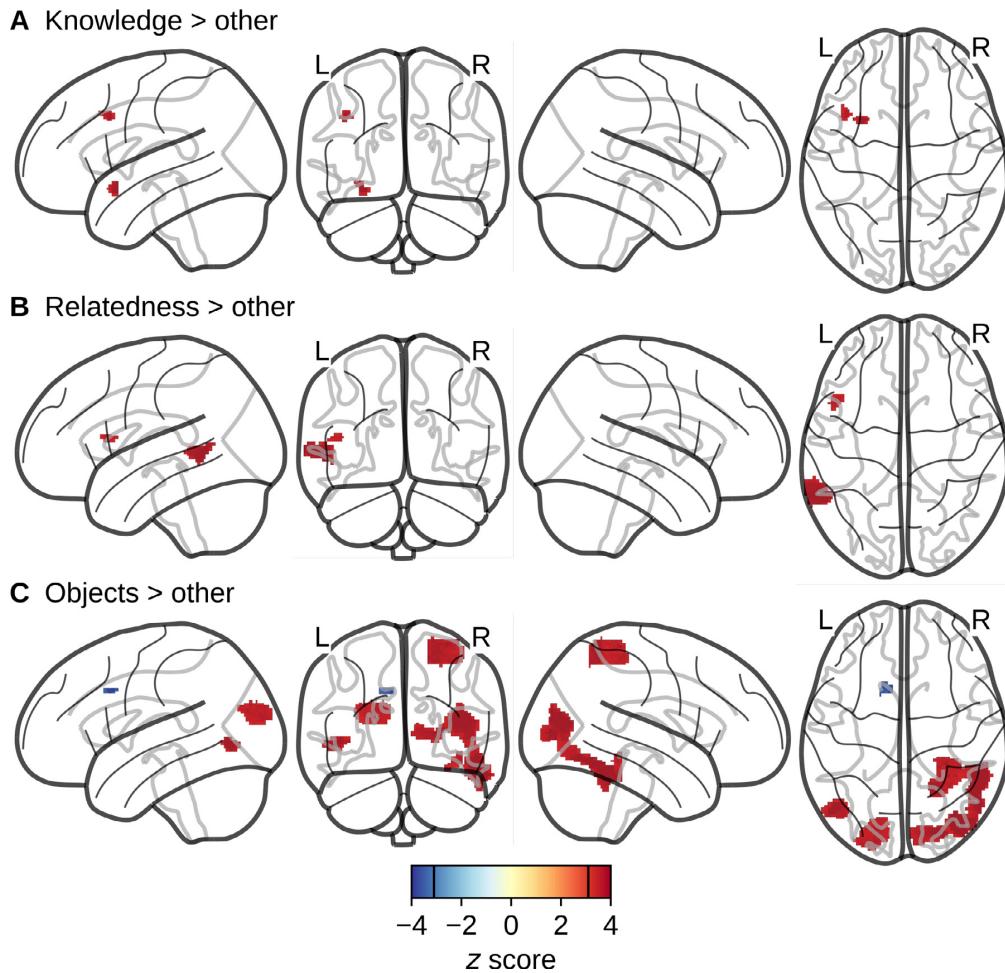


Fig. 6. Differences between semantic task categories.

For each task category shown in Figure 5, the meta-analytic ALE map was contrasted against the map for the experiments from the other two task categories. The resulting subtraction maps were thresholded at $p < .001$ (uncorrected) at the voxel level and $k > 25$ connected voxels (200 mm^3) at the cluster level.

tasks probing visual semantic object categories showed more consistent activation than the other two task categories in the bilateral occipital and fusiform cortices as well as in the right superior parietal cortex. They also showed reliably less activation than the other two task categories in one small cluster in the left medial frontal lobe (shown in blue in Fig. 6C). The conjunction analysis revealed that there were no regions of activation that were shared by all three semantic task categories (as seen also in Fig. 5B–D).

3.5. Age-related changes

Age-related changes in fMRI activation patterns across the 50 experiments were examined (a) by splitting the sample of experiments in half at the median of the (mean) sample ages ($md_{\text{mean age}} = 10.3$ years) and performing an ALE subtraction analysis as described above (see Section 2.4) and (b) by entering mean sample age as a linear predictor in a meta-regression model using the effect size-based SDM approach (see Section 2.3).

The median split-based approach showed more consistent activation in experiments with older (> 10.3 years) as compared to younger (< 10.3 years)

children only at the right putamen and insula (see Fig. 7A and Table 5). The effect size-based approach showed no age-related changes using the pre-specified statistical threshold ($p < .001$ [FWE-corrected] at the voxel level and $k > 25$ connected voxels [200 mm^3] at the cluster level).

However, one should consider that meta-analytic null effects for study-level moderating variables may in part reflect the lack of statistical power for detecting them (Hempel et al., 2013). This is especially true when the variable of interest (here: mean sample age) has a restricted variance (see Fig. 3 and Table 1). To mitigate this lack of statistical power in a post hoc fashion, we present the uncorrected and non-thresholded z score map from the effect size-based meta-regression in Fig. 7B. This map suggests an age-related decrease of effect sizes in the left middle/superior temporal gyrus (peak $z = -2.55$) and an age-related increase of effect sizes in the left inferior frontal gyrus (peak $z = 2.33$). To a lesser extent, the increase in the inferior frontal gyrus is mirrored in the right hemisphere, consistent with the median split-based result from ALE. However, none of these peaks survived our initial cluster-forming threshold and therefore additional experiments will be needed to confirm if these age-related changes turn out to be reliable on a meta-analytic level.

Table 4
Statistics for the meta-analytic clusters shown in Fig. 6.

Analysis	#	Size (mm ³)	Mean z	Peak z	Peak X	Peak Y	Peak Z	Peak anatomical label
Knowledge > (relatedness + objects)	1	304	3.62	4.06	-28	10	-16	L insula
	2	256	3.49	3.89	-40	14	36	L middle frontal gyrus
Relatedness > (knowledge + objects)	1	1960	3.72	4.06	-58	-50	0	L middle temporal gyrus
	2	336	3.42	3.62	-46	14	12	L inferior frontal gyrus (oper.)
Objects > (knowledge + relatedness + edge)	1	9912	3.71	4.06	46	-68	-4	R middle inferior gyrus
	2	4744	3.48	4.06	28	-46	66	R postcentral gyrus
	3	4048	3.45	3.89	-18	-84	26	L superior occipital gyrus
	4	1104	3.65	4.06	-46	-70	-2	L middle occipital gyrus
	5	312	3.48	4.06	38	-40	-14	R fusiform gyrus
	6	200	-3.52	-3.54	-10	12	34	L middle cingulate gyrus
Conjunction	<i>No significant clusters</i>							

Note. Peak anatomical labels are based on the anatomic automatic labeling atlas (AAL2; Rolls et al., 2015). # = cluster ID, L = left, R = right, oper. = pars opercularis.

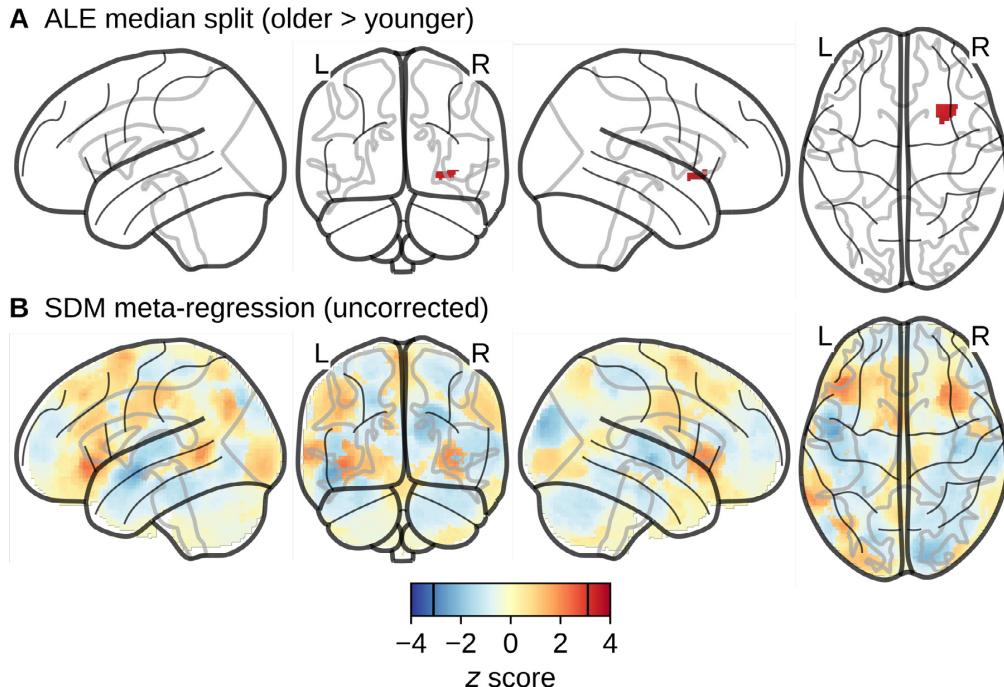


Fig. 7. Differences between older and younger children.

(A) The ALE map for experiments with a mean sample age older than the meta-analytic median (10.3 years) was contrasted against the ALE map for experiments with a mean sample age younger than this median. The resulting subtraction map was thresholded at $p < .001$ (uncorrected) at the voxel level and $k > 25$ connected voxels (200 mm^3) at the cluster level. (B) Meta-regression testing for a linear influence of mean sample age on meta-analytic effect sizes. Since none of the voxels met the prespecified statistical threshold ($p < .001$ [FWE-corrected] at the voxel level and $k > 25$ connected voxels [200 mm^3] at the cluster level), we present the uncorrected and non-thresholded z score map in an exploratory fashion and as a starting point for future research.

Table 5
Statistics for the meta-analytic clusters shown in Fig. 7.

Analysis	#	Size (mm ³)	Mean z	Peak z	Peak X	Peak Y	Peak Z	Peak anatomical label
Median split (older > younger)	1	536	3.41	3.62	28	12	-6	R putamen
Meta-regression	<i>No significant clusters</i>							

Note. Peak anatomical labels are based on the anatomic automatic labeling atlas (AAL2; Rolls et al., 2015). # = cluster ID, R = right.

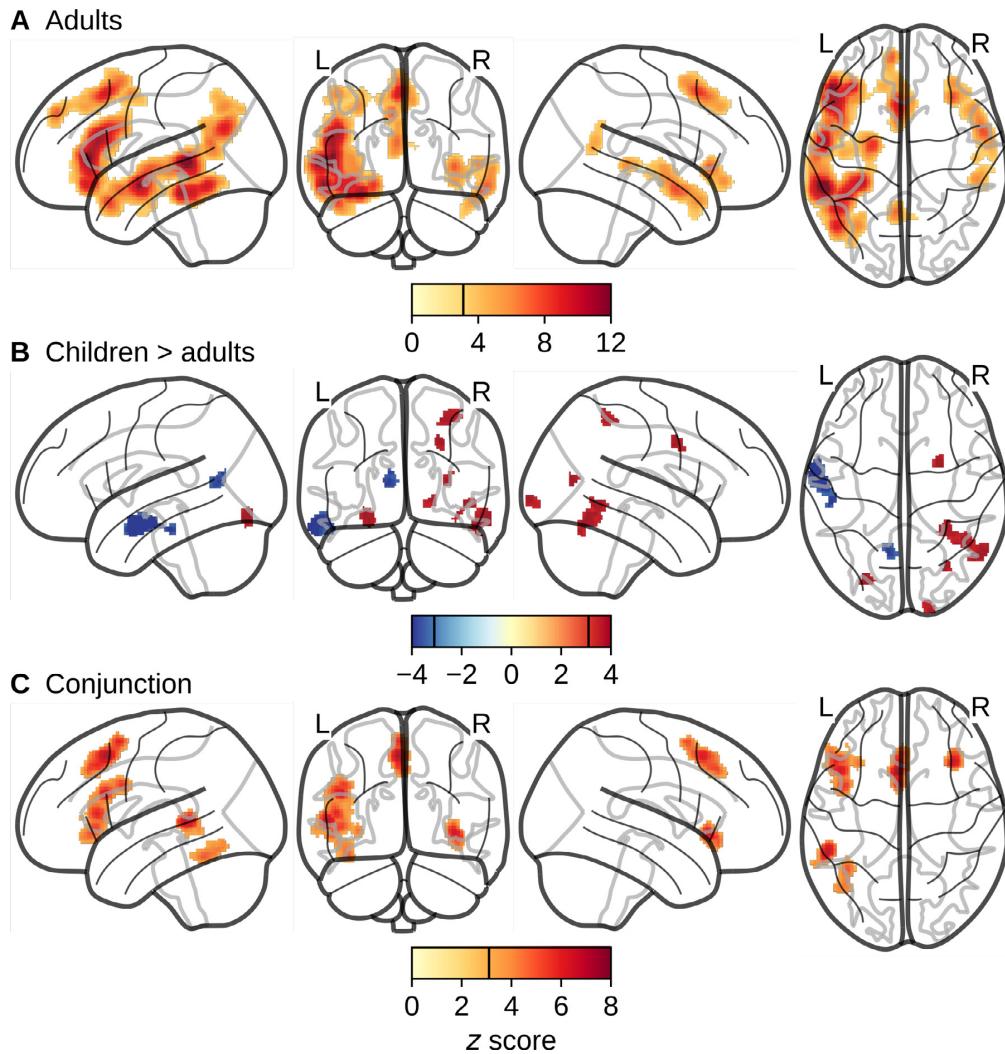


Fig. 8. Comparison with semantic cognition in adults.

(A) Reproduction of the meta-analysis by Jackson (2021), synthesizing 415 fMRI experiments of general semantic cognition in adults. To allow for a group comparison, this reproduction was created using the same data-analytic procedures and statistical thresholds as described in the main text and in Figure 4B for children. (B) Comparison between the ALE maps of semantic cognition in children and adults, thresholded at $p < .001$ (uncorrected) at the voxel level and $k > 25$ connected voxels (200 mm^3) at the cluster level. (C) Conjunction analysis showing only voxels that were significant for both children and adults. Here, the color indicates the minimum z score for each significant voxel across both individual group maps.

3.6. Comparison with semantic cognition in adults

A recent meta-analysis by Jackson (2021) used ALE to synthesize the fMRI literature on semantic control in adults. They also broadened their analysis to 415 studies of general semantic cognition and found wide-ranging clusters of consistent activation especially in the left hemisphere, spanning multiple areas in the temporal and inferior frontal lobes as well as the supplementary motor area (see Fig. 8A and Table 6 for a reproduction of these results based on the peak coordinates kindly provided by the original author). This meta-analytic map of semantic cognition in adults was compared to the map of semantic cognition in children by means of a subtraction and conjunction analysis. This revealed more consistent activation in children as compared to adults in multiple posterior regions of the cortex, including the bilateral inferior temporal and right inferior parietal gyri (see Fig. 8B and Table 6). In

contrast, adults showed more consistent activation in the anterior part of the left middle and inferior temporal gyri as well as deep in the left calcarine sulcus. Finally, the conjunction analysis indicated large areas of overlap between the two groups in the left inferior frontal gyrus, the supplementary motor area, the left middle and superior temporal gyri, the right insular and inferior frontal cortices, and the left fusiform gyrus (see Fig. 8C and Table 6).

3.7. Evaluation of robustness

The robustness of the meta-analytic results against two different types of publication bias

—spurious findings and the file drawer problem—was assessed using a leave-one-out (jackknife) analysis and a fail-safe N analysis. Both of these analyses were conducted for the entire sample of all 50 semantic

Table 6
Statistics for the meta-analytic clusters shown in Fig. 8.

Analysis	#	Size (mm ³)	Mean z	Mean ALE	Peak z	Peak ALE	Peak X	Peak Y	Peak Z	Peak anatomical label
Adults	1	88,648	5.41	0.064	13.27	0.179	-56	-38	2	L middle temporal gyrus
	2	12,488	4.65	0.054	9.68	0.121	-4	18	50	L supplementary motor area
	3	8320	4.35	0.050	7.54	0.090	56	0	-18	R middle temporal gyrus
	4	4176	4.35	0.050	7.37	0.087	36	24	-2	R insula
	5	3616	4.36	0.050	6.89	0.081	-6	-56	14	L precuneus
	6	1768	3.99	0.046	6.11	0.071	52	-34	0	R middle temporal gyrus
	7	72	3.22	0.038	3.35	0.039	-26	26	46	L middle frontal gyrus
> adults	1	1072	3.68		3.89		56	-64	-18	R inferior temporal gyrus
	2	776	3.59		3.89		30	-48	52	R inferior parietal gyrus
	3	672	3.59		3.48		44	-56	-8	R inferior temporal gyrus
	4	664	3.69		4.06		-26	-82	-16	L lingual gyrus
	5	432	3.51		4.06		20	-104	-2	R calcarine sulcus
	6	352	3.66		4.06		26	0	36	No label found
	7	232	3.69		3.39		32	-74	10	R middle occipital gyrus
Adults	1	2632	3.78		4.06		-58	-10	-22	L middle temporal gyrus
	2	584	3.49		4.06		-10	-62	10	L calcarine sulcus
Conjunction	1	9592	3.95	0.023	6.45	0.043	-44	18	24	L inferior frontal gyrus (tri.)
	2	5320	4.41	0.027	7.43	0.053	-4	16	50	L supplementary motor area
	3	2480	4.38	0.026	7.05	0.049	-52	-38	4	L middle temporal gyrus
	4	2232	4.23	0.028	6.74	0.053	36	24	-4	R insula
	5	2056	3.66	0.021	4.84	0.029	-40	-52	-20	L fusiform gyrus
	6	48	3.20	0.017	3.30	0.018	-40	36	6	L inferior frontal gyrus (tri.)

Note. Peak anatomical labels are based on the anatomic automatic labeling atlas (AAL2; [Rolls et al., 2015](#)). # = cluster ID, L = left, R = right, tri. = pars triangularis.

experiments (see Fig. 4B and Table 2) and for each of the three task category-specific sub-analyses (see Fig. 5B–D and Table 3).

The leave-one-out procedure showed that all clusters detected in the main analysis were robust against the deletion of individual studies, with an average leave-one-out robustness of 96% across the eight clusters (range 84–100%; see Fig. 9). For the sub-analysis of semantic relatedness experiments, the robustness of all four clusters was at 100%, whereas for semantic knowledge experiments, it was at 100% for Clusters #1 and #2 but reduced for Clusters #3 (right insula; 48%) and #4 (left IFG, 81%). Finally, for visual semantic object category experiments, it was slightly reduced (85%) for all three clusters. Together, this reflects good overall robustness against spurious experiments in the meta-analysis, although this robustness was compromised slightly for the sub-analyses that were run on fewer experiments (see also [Eickhoff et al., 2016](#)).

The fail-safe N analysis showed that most clusters were robust against the file drawer problem. This was indicated by the fact that in these cases, the number of (unpublished) null experiments that needed to be added until overturning the statistical significance of the cluster exceeded the number of (published) experiments in the original analysis (see Fig. 10). The only clusters were this was not the case were Clusters #5 (right occipital lobe; FSN = 18), #6 (right fusiform gyrus; FSN = 15), and #8 (left inferior frontal gyrus; FSN = 1). Note, however, that Clusters #5 and #6 still marginally exceeded the desired value of 30% of the original sample size (see Section 2.7 and [Samartsidis et al., 2020](#)). For the task category-specific sub-analyses, FSN values were high overall, except for the knowledge-related Clusters #3 (right insula; FSN = 1) and #4 (left inferior frontal gyrus; FSN = 2) as well as the object-related Clusters #1 (right fusiform gyrus; FSN = 8), #2 (left fusiform gyrus; FSN = 5), and #3 (right occipital lobe; FSN = 11). All but the two knowledge-related clusters exceeded the desired threshold of 30%, once more giving the overall impression of satisfactory robustness to publication bias. However, these two specific clusters need to be interpreted with caution and would require additional support by future fMRI experiments.

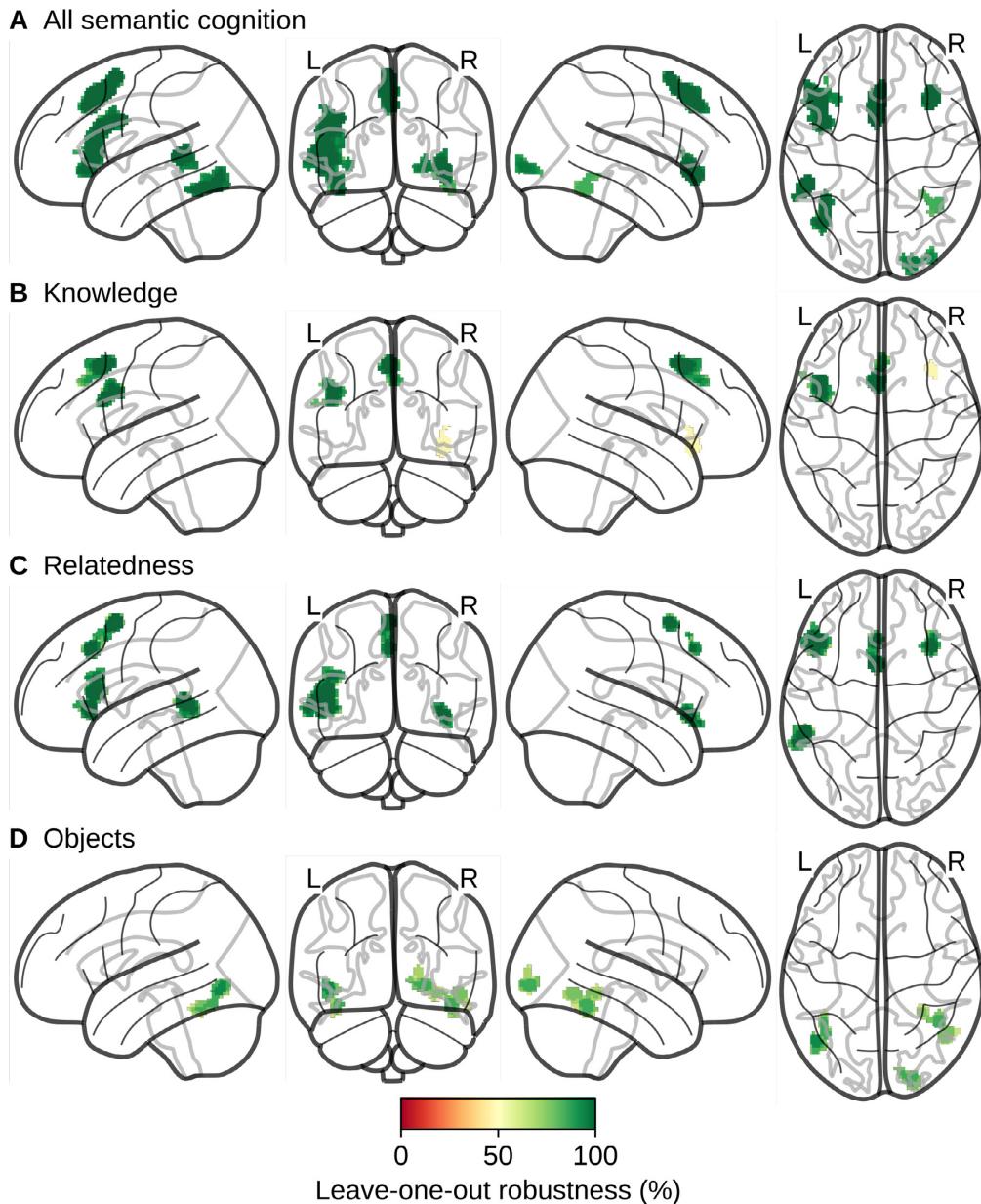
4. Discussion

Here we systematically localized the brain areas underlying semantic cognition in children by means of a coordinate-based meta-analysis of fMRI studies. We identified 50 individual experiments scanning chil-

dren with a mean age of 3–12 years using a variety of semantic tasks. Pooling across the reported peak coordinates from all of these experiments, we found evidence for consistent activation in sub-regions of the left perisylvian language network associated with lexical processing (left MTG/STG and IFG) as well as in the bilateral SMA, the right insula, and more posterior brain regions in the bilateral fusiform and right occipital cortices. These areas were recruited to a different degree by different semantic task categories: Inferior frontal regions and the SMA were recruited preferentially during tasks tapping into semantic knowledge (e.g., naming an object after hearing its descriptions) and semantic relatedness (e.g., hearing two words and deciding if they are related or not), while posterior regions were recruited preferentially during tasks tapping into the differentiation of visual object categories (e.g., passively viewing faces as compared to other visual stimuli).

The left MTG/STG and the pars triangularis of the left IFG are known to be implicated in semantic processing from at least 3 years of age onwards ([Skeide et al., 2014](#)). There is also some evidence that children are able to process word meaning in the left MTG with as little as 2 years of age ([Friedrich and Friederici, 2010](#); [Travis et al., 2011](#)). Within the left IFG, semantic processing especially recruits the more anterior parts (pars triangularis and pars orbitalis; [Brauer and Friederici, 2007](#); [Nuñez et al., 2011](#); [Skeide et al., 2014](#); [Skeide and Friederici, 2016](#)) which also showed the strongest meta-analytic peaks in our study. These sub-areas seem to play an especially crucial role in children's language processing, as they allow them to successfully retrieve the semantic meaning of grammatically challenging sentences even though their syntactic abilities (localized in the pars opercularis of the left IFG in adults) are not yet fully developed ([Skeide et al., 2014](#)).

Just as the left IFG and MTG/STG, the bilateral SMA also showed meta-analytically robust activation in children performing semantic knowledge and relatedness tasks. This mirrors previous meta-analyses of semantic cognition in adults (e.g., [Binder et al., 2009](#); [Jackson, 2021](#)) as well as meta-analyses of language comprehension in both children ([Enge et al., 2020](#); [Martin et al., 2015](#)) and adults (e.g., [Ferstl et al., 2008](#); [Rodd et al., 2015](#)). The premotor activation could reflect a grounding of abstract semantic concepts in articulatory motor representations ([Martin, 2016](#); [Pulvermüller and Fadiga, 2010](#)). Alternatively or in addition to this, activation in the anterior part of the SMA (pre-SMA) may also reflect higher-order cognitive control processes such as ambiguity resolution and the integration of semantic context ([Hertrich et al., 2016](#)). This is supported by our observation that the SMA showed

**Fig. 9.** Leave-one-out analysis.

For each meta-analytic ALE map (see Figures 4B and 5B–D), the original analysis was repeated as many times as there were experiments in the sample, each time leaving out another one of these experiments. The colors show the percentage of these simulations in which the cluster remained statistically significant and therefore invariant against the exclusion of any individual experiment (e.g., because the results may have been spurious).

no consistent activation when children performed visual object category tasks. These tasks would in many cases afford a similar degree of pre-motor response (e.g., when viewing tools; e.g., Dekker et al., 2014; Kersey et al., 2016) but arguably a lesser degree of cognitive control compared to tasks probing semantic knowledge or relatedness (see Binder et al., 2009, for similar findings focusing exclusively on experiments with linguistic stimuli in adults). Note, however, that only a limited number of visual object category tasks ($n = 13$) could be

included in the present meta-analysis, presumably limiting statistical power (Eickhoff et al., 2016).

Finally, the ventral temporal and occipital cortices (fusiform gyrus and adjacent areas) are well-known to house category-selective neuronal populations that respond primarily to certain categories of visual stimuli (e.g., faces in the fusiform face area [FFA] or objects in the lateral occipital complex [LOC]; Grill-Spector and Weiner, 2014; Haxby et al., 2001). Accordingly, these regions showed consistent activation only for

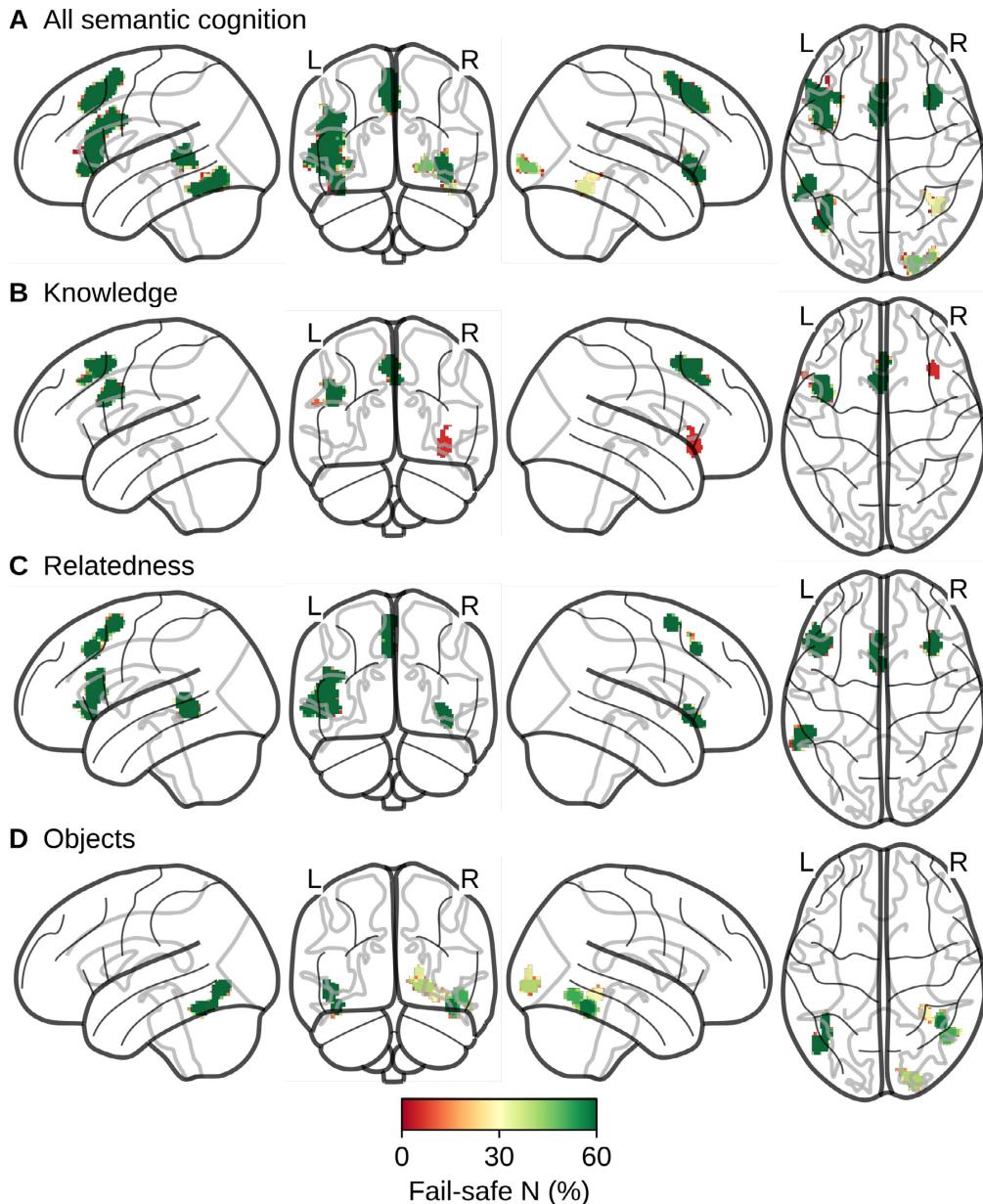


Fig. 10. Fail-safe N analysis.

For each meta-analytic ALE map (see Figures 4B and 5B–D), the original analysis (with n experiments) was repeated up to $5n$ times, each time adding one additional null experiment with peaks of activation distributed randomly within the gray matter mask. The FSN metric for each cluster was computed as the highest number of null experiments that could be added so that the cluster still remained statistically significant. To accommodate for the different sample sizes of the four subanalyses, the FSN is shown as the percentage of the number of original experiments. Values greater than 30% exceed the most conservative estimate for the actual size of the file drawer problem in the fMRI literature (Samartsidis et al., 2020).

tasks in which children viewed different visual semantic object categories. A subset of small areas within this ventral temporal and occipital region also showed the most prominent increase of meta-analytic activation for children as compared to adults. This may reflect that children need to recruit these patches of cortex to a stronger degree than adults to distinguish between different kinds of visual stimuli (see also Antonucci and Alt, 2011). Most of these child-specific activations were

observed in the right hemisphere. Therefore, it seems as though the semantic network of most children had not yet reached its adult-like level of lateralization. This ongoing process of semantic lateralization seems closely connected to lateralization processes of the language comprehension network, some of which also seem to last until early adolescence (Berl et al., 2014; Enge et al., 2020; Holland et al., 2007). We cannot preclude, however, that these group differences could also be driven

by differences in the kinds of tasks and baseline conditions chosen to investigate semantic cognition in children and adults. Furthermore, the semantic categorization of different kinds of visual objects is confounded with lower-level sensory differences between them. These visual confounds might at least partially explain our meta-analytic results for this task category (e.g., the activation of the right early visual cortex; see Fig. 5D).

There was one region, namely the left ATL, that is oftentimes considered to be at the core of the semantic system (Lambon Ralph et al., 2017; Patterson et al., 2007) but did not show any meta-analytically consistent activation in children whatsoever. In adults, the ATL serves as an amodal “hub” connecting different modality-specific sites within the wider semantic network (e.g., speech processing in the IFG and visual semantics in the occipital and ventral temporal cortices). Neuroimaging studies of semantic cognition in children seem to elicit significantly less of such ATL activation (see Fig. 8B), suggesting that this semantic hub may need time to develop over childhood and into adolescence (Hwang et al., 2013, but also see Stevens et al., 2009). In contrast, most other regions of the semantic network showed at least some overlap between children and adults, especially in the left IFG, bilateral SMA, right insula, left MTG/STG, and left FG (see Fig. 8C and Jackson, 2021).

Because meta-analyses depend critically on the quality of the underlying literature, they can be prone to a number of biases (e.g., positivity bias and selective reporting). Tools to detect the presence of such biases are less well developed in meta-analytic frameworks for neuroimaging as compared to clinical or behavioral outcomes (Acar et al., 2018). However, in the present study, the number of reported peak coordinates—a very rough analogue of an experiment-specific effect size for fMRI studies—was unrelated to the sample size of the experiments (see Fig. 3). This is consistent with small sample bias and/or selective reporting in larger studies. To assess the robustness of our meta-analytic results against these kinds of biases, we first conducted a leave-one-out analysis which showed that all clusters were considerably invariant against the deletion of individual experiments from the sample. This means that we would have obtained identical results even if any of the original experiments reported only spurious activations. Second, we also conducted a fail-safe N analysis in which we estimated the number of unpublished null experiments (i.e., experiments without any consistent pattern of activation) that have to be added until the significance of any observed cluster is overturned. This number was larger than 30% of the meta-analytic sample size for almost all clusters. It thereby exceeded the current most conservative estimate for the actual number of unpublished fMRI experiments that are hidden in the “file drawer” (Samartsidis et al., 2020). Thus, although our meta-analysis could not directly assess or correct for publication bias in the underlying literature, its results seem to be stable even if one accepts that such biases are present.

5. Conclusion

Conducting fMRI experiments with children is challenging and costly, which is why sample sizes are often lower than in behavioral

experiments or in neuroimaging experiments with adult participants. Meta-analyses are therefore necessary to filter out spurious results and to uncover similarities and differences between different task paradigms. Regarding children’s capacity to process semantic information, our coordinate-based meta-analysis showed reliable patterns of activation in the left IFG and MTG/STG, the bilateral SMA, the right insula and parts of the bilateral ventral temporal and occipital cortices. Within this network, tasks probing children’s semantic world knowledge and semantic relatedness between stimuli showed overlapping spots of activation that were distinct from those seen in tasks probing the differentiation of visual semantic object categories. A comparison to the adult semantic system revealed largely overlapping regions of activation but also more child-specific activation in bilateral inferior temporal and occipital regions as well as more adult-specific activation in the anterior portion of the left temporal lobe.

Declaration of Competing Interest

None.

Credit authorship contribution statement

Alexander Enge: Conceptualization, Methodology, Software, Validation, Formal analysis. **Rasha Abdel Rahman:** Conceptualization, Writing – review & editing. **Michael A. Skeide:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration.

Acknowledgements

We would like to thank the developers of the NiMARE software package (especially Taylor Salo and Tal Yarkoni) for their assistance as well as Rebecca L. Jackson, Timothy A. Keller, Saloni Krishnan, and Robin J. Schafer for providing us with additional data that were not publicly available. The comments by two anonymous reviewers have helped us to improve an earlier version of this paper. This work was supported by a research grant (433715509) and a Heisenberg grant (433758790) of the German Research Foundation (DFG) awarded to M.A.S.

Appendix A

Additional information about the experiments included in the meta-analysis

#	Group identifier	Language of experiment	Modality of presentation	Modality of response	Soft-ware	Field strength (T)	FHWM (mm ³)	Voxelthr.	Cluster thr.	Peak table	Peak space	Peak stat.
1	Chronological age-matched (CA/NC)	English	Visual	Manual	AFNI	3	4	p < .005 (uncorr.)	p < .05; k > 400	Table 5: Categorize: NC	TAL	NA
2	Reading-matched (RM)	English	Visual	Manual	AFNI	3	4	p < .005 (uncorr.)	p < .05; k > 400	Table 5: Categorize: RM	TAL	NA
3	Younger children	English	Visual	None	MEDx	1.5	4	z > 2.4 (uncorr.)	p < .05	Table 2: Younger Group	TAL	NA
4	Normals	Dutch	Visual	Manual	SPM	1.5	8	p < .001 (uncorr.)	p < .05	Table 1: Semantic judgment: Normals	TAL	z
5	Only one group	English	Auditory	Covert	SPM	1.5	8	p < .001 (corr.)	k > 10	Table 1	TAL	z
6	Only one group	English	Auditory	Manual	SPM	1.5	8	p < .05 (corr.)	NA	Table 3	TAL	z
7	Children	English	Visual	Covert	SPM	3	6	p < .005 (uncorr.)	p < .05; k > 42	Table 2: Child group; Semantic retrieval main effect	MNI	t
8	Only one group	English	Auditory	Manual	SPM	3	8	p < .05 (corr.)	NA	Table II: Group map activation across all ages 4-12	MNI	t
9	Children	English	Auditory/visual	Manual	SPM	1.5	7	z > 4.5 (uncorr.)	k > 12	Appendix A / B: Children: Semantic	TAL	z
10	Children	English	Auditory/visual	Manual	SPM	1.5	7	p < .001 (uncorr.)	k > 15	Table 4: Meaning	MNI	z
11	Controls	English	Auditory/visual	Manual	SPM	1.5	10	p < .001 (uncorr.)	k > 15	Table 3 / 4: Controls	MNI	z
12	Children	German	Auditory	Manual	LIPSIA	3	4.239	z > 3.09 (uncorr.)	k > 10	Table 3: SEM vs. Baseline	TAL	z
13	Children	Chinese	Visual	Manual	SPM	2	7	p < .001 (uncorr.)	k > 20	Table IV: Children (meaning)	MNI	z
14	Children	English	Visual	Manual	SPM	1.5	10	p < .001 (uncorr.)	k > 14	Table II	TAL	z
15	Only one group	English	Auditory	Manual	SPM	1.5	10	p < .05 (corr.)	NA	Table 2	TAL	z
16	Only one group	Chinese	Visual	Manual	SPM	3	10	p < .05 (corr.)	k > 10	Table 2	MNI	z
17	Only one group	Chinese	Visual	Manual	SPM	3	10	p < .001 (uncorr.)	NA	Table 2: Time 1	MNI	z
18	Typical	English	Visual	Manual	SPM	1.5	4	p < .001 (uncorr.)	NA	Table 4: Person > Control: Typical	MNI	NA
19	9 to 10	English	Visual	Manual	FSL	1.5	5	z > 2.3 (uncorr.)	p < .05	Table A2: 9-10	MNI	z
20	7 to 8	English	Visual	Manual	FSL	1.5	5	z > 2.3 (uncorr.)	p < .05	Table A2: 7-8	MNI	z
21	Children	Chinese	Visual	Manual	SPM	3	8	p < .05 (corr.)	k > 10	Supplementary	MNI	z
22	Only one group	English	Visual	Covert	SPM	1.5	NA	z > 3.09 (corr.)	p < .05	Table 1: Contrast 1	TAL	z
23	Children	English	Visual	Covert	SPM	1.5	8	p < .0001 (corr.)	NA	Table 1: Children	TAL	z
24	Only one group	English	Audiovisual	Manual	FSL	3	8	z > 2.3 (corr.)	NA	Table 2	MNI	NA
25	Children	English	Visual	Manual	Brain-Voyager	3	6	p < .005 (uncorr.)	NA	Table 1: Children	TAL	NA
26	Children	English	Visual	Overt	FSL	1.5	6	z > 3.1 (uncorr.)	p < .05	Obtained from the authors	MNI	z
27	Typically developing	English	Visual	Overt	FSL	3	5	z > 6 (uncorr.)	k > 50	Table 3: A. TD	MNI	z
28	Visual-auditory judgment group	Chinese	Audiovisual	Manual	SPM	3	10	p < .05 (corr.)	k > 10	Table 2: Visual-auditory	MNI	z
29	Visual-visual judgment group	Chinese	Visual	Manual	SPM	3	10	p < .05 (corr.)	k > 10	Table 2: Visual-visual	MNI	z

(continued on next page)

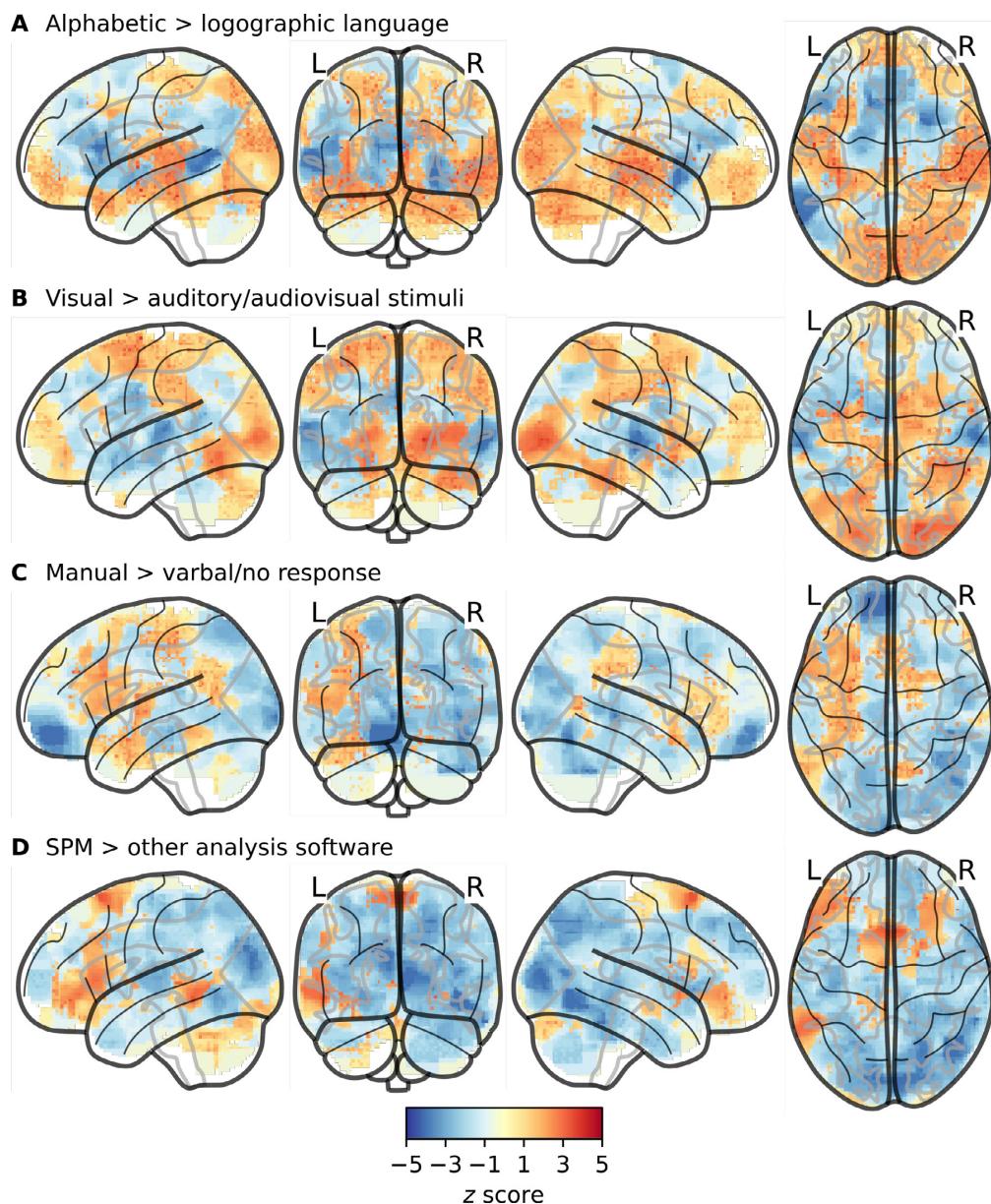
(continued)

#	Group identifier	Language of experiment	Modality of presentation	Modality of response	Soft-ware	Field strength (T)	FHWM (mm ³)	Voxelthr.	Cluster thr.	Peak table	Peak space	Peak stat.
30	Conjunction of children andadol.	Chinese	Visual	Manual	SPM	3	6	<i>p</i> < .05 (corr.)	<i>k</i> > 10	Table 2	MNI	<i>z</i>
31	Children	English	Visual	Manual	SPM	3	8	<i>p</i> < .01 (uncorr.)	<i>p</i> < .05; <i>k</i> > 8	Table 3: Children: <i>F</i> > (D and L) Table A1: SEMCAT	MNI	<i>z</i>
32	Only one group	German	Visual	Manual	SPM	3	8	<i>p</i> < .001 (uncorr.)	<i>p</i> < .05; <i>k</i> > 50	Table 3	MNI	<i>t</i>
33	Only one group	English	Visual	Manual	SPM	3	8	NA	<i>p</i> < .05; <i>k</i> > 20	Table 3	MNI	<i>z</i>
34	Good readers	English	Visual	Manual	SPM	3	8	<i>p</i> < .002 (uncorr.)	<i>k</i> > 10	Obtained from the authors	MNI	<i>t</i>
35	Normal readers	French	Visual	Manual	SPM	3	5	<i>p</i> < .001 (uncorr.)	<i>p</i> < .05	Table 2: Normal readers: Faces > others / Houses > others	MNI	<i>z</i>
36	Children	English	Auditory	Manual	AFNI	3	6	<i>t</i> > 3.786 (uncorr.)	<i>p</i> < .05; <i>k</i> > 208	Table 2	TAL	<i>t</i>
37	Typically developing children	Japanese	Visual	Manual	SPM	1.5	8	<i>p</i> < .001 (uncorr.)	<i>p</i> < .05	Table 4: TD children	MNI	<i>z</i>
38	Children	English	Visual	Manual	AFNI	1.5	NA	<i>p</i> < .001 (uncorr.)	<i>k</i> > 177	Table 1: Children: Face task	TAL	NA
39	Term-born children	English	Visual	Manual	SPM	1.5	8	<i>t</i> > 1.01 (uncorr.)	<i>p</i> < .01; <i>k</i> > 102	Obtained from the authors	TAL	<i>t</i>
40	Young adolescents	English	Visual	None	Brain-Voyager	3	0	<i>t</i> > 2.5 (uncorr.)	<i>p</i> < .05; <i>k</i> > 190	Table 2: Adolescents	TAL	NA
41	Children	English	Visual	None	Brain-Voyager	3	0	<i>t</i> > 2.5 (uncorr.)	<i>p</i> < .05; <i>k</i> > 190	Table 2: Children	TAL	NA
42	Typically developing	English	Visual	None	Brain-Voyager	3	0	<i>t</i> > 2.3 (uncorr.)	<i>p</i> < .05; <i>k</i> > 4	Table S2: TD adolescents	TAL	NA
43	Normal readers	Chinese	Visual	Manual	SPM	2	6	<i>p</i> < .05 (corr.)	<i>k</i> > 20	Table 1: Normal readers	MNI	<i>t</i>
44	9-to-10-year-olds	German	Audiovisual	Manual	SPM	3	4	<i>p</i> < .01 (corr.)	<i>p</i> < .05; <i>k</i> > 17	Table S3: C: Main effect semantic implausibility	MNI	<i>z</i>
45	6-to-7-year-olds	German	Audiovisual	Manual	SPM	3	4	<i>p</i> < .01 (corr.)	<i>p</i> < .05; <i>k</i> > 17	Table S3: B: Main effect semantic implausibility	MNI	<i>z</i>
46	Only one group	English	Auditory	Covert	NA	3	6	<i>z</i> > 6 (uncorr.)	<i>p</i> < .05; <i>k</i> > 10	Table 2: Localization of BOLD signal changes in all healthy subjects	TAL	NA
47	Controls	English	Auditory	Manual	ITI	3	NA	<i>z</i> > 7 (uncorr.)	<i>p</i> < .05; <i>k</i> > 30	Table 3: A: Semantic decision task: Group activation maps; Controls	TAL	NA
48	Typically developing youths	Chinese	Visual	Manual	SPM	3	10	<i>p</i> < .001 (uncorr.)	<i>k</i> > 10	Table S1: TD group	MNI	<i>z</i>
49	Children	German	Auditory	None	SPM	3	6	<i>p</i> < .005 (uncorr.)	<i>p</i> < .05; <i>k</i> > 27	Table 1: Five-year-old children: (B) Main effect of animacy hierarchy	MNI	<i>z</i>
50	Only one group	Chinese	Visual	Manual	SPM	2	8	<i>p</i> < .0001 (uncorr.)	<i>k</i> > 5	Table 1: Chinese minus baseline	MNI	<i>z</i>

Note. For articles reporting fMRI data from multiple groups (e.g., children and adults or typically developing children and neurodiverse children), the group-identifier indicates which group(s) were included as experiments in the meta-analysis. Please refer to Table 1 and/or the original research articles for further information about these groups of children. # = experiment ID, FHWM = smoothing kernel full width at half maximum, thr. = threshold, stat. = type of test statistic reported for individual peaks, uncorr. = not corrected for multiple comparisons, corr. = corrected for multiple comparisons, TAL = Talairach space, MNI = Montreal Neurological Institute space, NA = not available.

Appendix B

Fig. B1

**Fig. B1.** Effects of four different experiment-level covariates.

Meta-analytic subtraction analyses were carried out to compare pairs of activation likelihood maps as described in Section 2.4 of the main text. No statistical threshold was applied so as to highlight the exploratory nature of this analysis. (A) Experiments conducted in alphabetic languages (German, Dutch, French, English) as compared to logographic languages (Japanese, Mandarin Chinese), (B) experiments with visual as compared to auditory/audiovisual presentation of stimuli, (C) experiments with manual (button-press) as compared to verbal (covert/overt) or no response tasks, (D) experiments which used SPM as compared to any other software packages for fMRI analysis.

References

- Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G., 2014. Machine learning for neuroimaging with scikit-learn. *Front Neuroinform* 8, 1–10. doi:[10.3389/fninf.2014.00014](https://doi.org/10.3389/fninf.2014.00014).
- Acar, F., Seurinck, R., Eickhoff, S.B., Moerkerke, B., 2018. Assessing robustness against potential publication bias in activation likelihood estimation (ALE) meta-analyses for fMRI. *PLoS One* 13 (11), e0208177. doi:[10.1371/journal.pone.0208177](https://doi.org/10.1371/journal.pone.0208177).
- Albaiges-Eizagirre, A., Solanes, A., Radua, J., 2019a. Meta-analysis of non-statistically significant unreported effects. *Stat. Methods Med. Res.* 28 (12), 3741–3754. doi:[10.1177/0962280218811349](https://doi.org/10.1177/0962280218811349).
- Albaiges-Eizagirre, A., Solanes, A., Vieta, E., Radua, J., 2019b. Voxel-based meta-analysis via permutation of subject images (PSI): theory and implementation for SDM. *NeuroImage* 186, 174–184. doi:[10.1016/j.neuroimage.2018.10.077](https://doi.org/10.1016/j.neuroimage.2018.10.077).
- Antonucci, S.M., Alt, M., 2011. A lifespan perspective on semantic processing of concrete concepts: does a sensory/motor model have the potential to bridge the gap? *Cognit. Affect. Behav. Neurosci.* 11 (4), 551–572. doi:[10.3758/s13415-011-0053-y](https://doi.org/10.3758/s13415-011-0053-y).
- Arnoldussen, A.L., 2006. *Neural Systems Involved in reading: FMRI Studies of orthography, Phonology and Semantics* (Order No. 3234758) [Doctoral dissertation, University of Wisconsin-Madison]. ProQuest Dissertations and Theses Global. <https://search.proquest.com/docview/304976526?accountid=11531>.
- Aylward, E.H., Park, J.E., Field, K.M., Parsons, A.C., Richards, T.L., Cramer, S.C., Meltzoff, A.N., 2005. Brain activation during face perception: evidence of a developmental change. *J. Cogn. Neurosci.* 17 (2), 308–319. doi:[10.1162/0898929053124884](https://doi.org/10.1162/0898929053124884).
- Backes, W., Vuurman, E., Wenckens, R., Spronk, P., Wuisman, M., van Engelschoven, J., Jolles, J., 2002. Atypical brain activation of reading processes in children with developmental dyslexia. *J. Child Neurol.* 17 (12), 867–871. doi:[10.1177/08830738020170121601](https://doi.org/10.1177/08830738020170121601).
- Balsamo, L.M., Xu, B., Gaillard, W.D., 2006. Language lateralization and the role of the fusiform gyrus in semantic processing in young children. *NeuroImage* 31 (3), 1306–1314. doi:[10.1016/j.neuroimage.2006.01.027](https://doi.org/10.1016/j.neuroimage.2006.01.027).
- Balsamo, L.M., Xu, B., Grandin, C.B., Petrella, J.R., Braniecki, S.H., Elliott, T.K., Gaillard, W.D., 2002. A functional magnetic resonance imaging study of left hemisphere language dominance in children. *Arch. Neurol.* 59 (7), 1168–1174. doi:[10.1001/archneur.59.7.1168](https://doi.org/10.1001/archneur.59.7.1168).
- Barquier, L.A., Davis, N., Cutting, L.E., 2014. Neuroimaging of reading intervention: a systematic review and activation likelihood estimate meta-analysis. *PLoS One* 9 (1), e83668. doi:[10.1371/journal.pone.0083668](https://doi.org/10.1371/journal.pone.0083668).
- Bauer, P.J., Pathman, T., Inman, C., Campanella, C., Hamann, S., 2017. Neural correlates of autobiographical memory retrieval in children and adults. *Memory* 25 (4), 450–466. doi:[10.1080/09658211.2016.1186699](https://doi.org/10.1080/09658211.2016.1186699).
- Berl, M.M., Mayo, J., Parks, E.N., Rosenberger, L.R., VanMeter, J., Ratner, N.B., Vaidya, C.J., Gaillard, W.D., 2014. Regional differences in the developmental trajectory of lateralization of the language network. *Hum. Brain Mapp.* 35 (1), 270–284. doi:[10.1002/hbm.22179](https://doi.org/10.1002/hbm.22179).
- Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cereb. Cortex* 19 (12), 2767–2796. doi:[10.1093/cercor/khp055](https://doi.org/10.1093/cercor/khp055).
- Booth, J.R., Burman, D.D., Van Santen, F., Harasaki, Y., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., 2001. The development of specialized brain systems in reading and oral-language. *Child Neuropsychol.* 7 (3), 119–141. doi:[10.1076/chin.7.3.119.8740](https://doi.org/10.1076/chin.7.3.119.8740).
- Booth, J.R., Bebko, G., Burman, D.D., Bitan, T., 2007. Children with reading disorder show modality independent brain abnormalities during semantic tasks. *Neuropsychologia* 45 (4), 775–783. doi:[10.1016/j.neuropsychologia.2006.08.015](https://doi.org/10.1016/j.neuropsychologia.2006.08.015).
- Booth, J.R., Burman, D.D., Meyer, J.R., Lei, Z., Choy, J., Gitelman, D.R., Parrish, T.B., Mesulam, M.M., 2003. Modality-specific and -independent developmental differences in the neural substrate for lexical processing. *J. Neurolinguistics* 16 (4–5), 383–405. doi:[10.1016/S0911-6044\(03\)00019-8](https://doi.org/10.1016/S0911-6044(03)00019-8).
- Brauer, J., Friederici, A.D., 2007. Functional neural networks of semantic and syntactic processes in the developing brain. *J. Cogn. Neurosci.* 19 (10), 1609–1623. doi:[10.1162/jocn.2007.19.10.1609](https://doi.org/10.1162/jocn.2007.19.10.1609).
- Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafo, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* 14 (5), 365–376. doi:[10.1038/nrn3475](https://doi.org/10.1038/nrn3475).
- Cao, F., Peng, D., Liu, L., Jin, Z., Fan, N., Deng, Y., Booth, J.R., 2008. Developmental differences of neurocognitive networks for phonological and semantic processing in Chinese word reading. *Hum. Brain Mapp.* 30 (3), 797–809. doi:[10.1002/hbm.20546](https://doi.org/10.1002/hbm.20546).
- Chou, T.-L., Booth, J.R., Bitan, T., Burman, D.D., Bigio, J.D., Cone, N.E., Lu, D., Cao, F., 2006a. Developmental and skill effects on the neural correlates of semantic processing to visually presented words. *Hum. Brain Mapp.* 27 (11), 915–924. doi:[10.1002/hbm.20231](https://doi.org/10.1002/hbm.20231).
- Chou, T.-L., Booth, J.R., Burman, D.D., Bitan, T., Bigio, J.D., Lu, D., Cone, N.E., 2006b. Developmental changes in the neural correlates of semantic processing. *NeuroImage* 29 (4), 1141–1149. doi:[10.1016/j.neuroimage.2005.09.064](https://doi.org/10.1016/j.neuroimage.2005.09.064).
- Chou, T.-L., Chen, C.-W., Fan, L.-Y., Chen, S.-Y., Booth, J.R., 2009. Testing for a cultural influence on reading for meaning in the developing brain: the neural basis of semantic processing in Chinese children. *Front. Hum. Neurosci.* 3, 27. doi:[10.3389/neuro.09.027.2009](https://doi.org/10.3389/neuro.09.027.2009).
- Chou, T.-L., Wong, C.-H., Chen, S.-Y., Fan, L.-Y., Booth, J.R., 2019. Developmental changes of association strength and categorical relatedness on semantic processing in the brain. *Brain Lang.* 189, 10–19. doi:[10.1016/j.bandl.2018.12.006](https://doi.org/10.1016/j.bandl.2018.12.006).
- Cocquyt, E.M., Coffé, C., van Mierlo, P., Duyck, W., Mariën, P., Szmałec, A., Santens, P., De Letter, M., 2019. The involvement of subcortical grey matter in verbal semantic comprehension: a systematic review and meta-analysis of fMRI and PET studies. *J. Neurolinguistics* 51, 278–296. doi:[10.1016/j.jneuroling.2019.04.001](https://doi.org/10.1016/j.jneuroling.2019.04.001).
- Corbett, B.A., Carmean, V., Ravizza, S., Wendelken, C., Henry, M.L., Carter, C., Rivera, S.M., 2009. A functional and structural study of emotion and face processing in children with autism. *Psychiatry Res.* 173 (3), 196–205. doi:[10.1016/j.psychres.2008.08.005](https://doi.org/10.1016/j.psychres.2008.08.005).
- Dekker, T.M., Mareschal, D., Johnson, M.H., Sereno, M.I., 2014. Picturing words? sensorimotor cortex activation for printed words in child and adult readers. *Brain Lang.* 139, 58–67. doi:[10.1016/j.bandl.2014.09.009](https://doi.org/10.1016/j.bandl.2014.09.009).
- Deniz, F., Nunez-Elizalde, A.O., Huth, A.G., Gallant, J.L., 2019. The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *J. Neurosci.* 39 (39), 7722–7736. doi:[10.1523/JNEUROSCI.0675-19.2019](https://doi.org/10.1523/JNEUROSCI.0675-19.2019).
- Eickhoff, S.B., Bzdok, D., Laird, A.R., Kurth, F., Fox, P.T., 2012. Activation likelihood estimation meta-analysis revisited. *NeuroImage* 59 (3), 2349–2361. doi:[10.1016/j.neuroimage.2011.09.017](https://doi.org/10.1016/j.neuroimage.2011.09.017).
- Eickhoff, S.B., Laird, A.R., Grefkes, C., Wang, L.E., Zilles, K., Fox, P.T., 2009. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty. *Hum. Brain Mapp.* 30 (9), 2907–2926. doi:[10.1002/hbm.20718](https://doi.org/10.1002/hbm.20718).
- Eickhoff, S.B., Nichols, T.E., Laird, A.R., Hoffstaedter, F., Amunts, K., Fox, P.T., Bzdok, D., Eickhoff, C.R., 2016. Behavior, sensitivity, and power of activation likelihood estimation characterized by massive empirical simulation. *NeuroImage* 137, 70–85. doi:[10.1016/j.neuroimage.2016.04.072](https://doi.org/10.1016/j.neuroimage.2016.04.072).
- Enge, A., Friederici, A.D., Skeide, M.A., 2020. A meta-analysis of fMRI studies of language comprehension in children. *NeuroImage* 215, 116858. doi:[10.1016/j.neuroimage.2020.116858](https://doi.org/10.1016/j.neuroimage.2020.116858).
- Fairhall, S.L., Caramazza, A., 2013. Brain regions that represent amodal conceptual knowledge. *J. Neurosci.* 33 (25), 10552–10558. doi:[10.1523/JNEUROSCI.0051-13.2013](https://doi.org/10.1523/JNEUROSCI.0051-13.2013).
- Fan, L.-Y., Lo, Y.-C., Hsu, Y.-J., Chen, W.-Y.I., Chou, T.-L., 2020. Developmental differences of structural connectivity and effective connectivity in semantic judgments of Chinese characters. *Front. Hum. Neurosci.* 14, 233. doi:[10.3389/fnhum.2020.00223](https://doi.org/10.3389/fnhum.2020.00223).
- Forstl, E.C., Neumann, J., Bogler, C., von Cramon, D.Y., 2008. The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Hum. Brain Mapp.* 29 (5), 581–593. doi:[10.1002/hbm.20422](https://doi.org/10.1002/hbm.20422).
- Fonov, V., Evans, A.C., Botteron, K., Almlöf, C.R., McKinstry, R.C., Collins, D.L., 2011. Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage* 54 (1), 313–327. doi:[10.1016/j.neuroimage.2010.07.033](https://doi.org/10.1016/j.neuroimage.2010.07.033).
- Friedrich, M., Friederici, A.D., 2010. Maturing brain mechanisms and developing behavioral language skills. *Brain Lang.* 114 (2), 66–71. doi:[10.1016/j.bandl.2009.07.004](https://doi.org/10.1016/j.bandl.2009.07.004).
- Gaillard, W.D., Pugliese, M., Grandin, C.B., Braniecki, S.H., Kondapaneni, P., Hunter, K., Xu, B., Petrella, J.R., Balsamo, L., Bassi, G., 2001. Cortical localization of reading in normal children: an fMRI language study. *Neurology* 57 (1), 47–54. doi:[10.1212/WNL.57.1.47](https://doi.org/10.1212/WNL.57.1.47).
- Gaillard, W.D., Sachs, B.C., Whitnah, J.R., Ahmad, Z., Balsamo, L.M., Petrella, J.R., Braniecki, S.H., McKinney, C.M., Hunter, K., Xu, B., Grandin, C.B., 2003. Developmental aspects of language processing: fMRI of verbal fluency in children and adults. *Hum. Brain Mapp.* 18 (3), 176–185. doi:[10.1002/hbm.10091](https://doi.org/10.1002/hbm.10091).
- Grill-Spector, K., Weiner, K.S., 2014. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat. Rev. Neurosci.* 15 (8), 536–548. doi:[10.1038/nrn3747](https://doi.org/10.1038/nrn3747).
- Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P., 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293 (5539), 2425–2430. doi:[10.1126/science.1063736](https://doi.org/10.1126/science.1063736).
- Hempel, S., Miles, J.N., Booth, M.J., Wang, Z., Morton, S.C., Shekelle, P.G., 2013. Risk of bias: a simulation study of power to detect study-level moderator effects in meta-analysis. *Syst. Rev.* 2, 107. doi:[10.1186/2046-4053-2-107](https://doi.org/10.1186/2046-4053-2-107).
- Hertrich, I., Dietrich, S., Ackermann, H., 2016. The role of the supplementary motor area for speech and language processing. *Neurosci. Biobehav. Rev.* 68, 602–610. doi:[10.1016/j.neubiorev.2016.06.030](https://doi.org/10.1016/j.neubiorev.2016.06.030).
- Holland, S.K., Vannest, J., Mecoli, M., Jacala, L.M., Tillemans, J.-M., Karunaratna, P.R., Schmitzorst, V.J., Yuan, W., Plante, E., Byars, A.W., 2007. Functional MRI of language lateralization during development in children. *Int. J. Audiol.* 46 (9), 533–551. doi:[10.1080/14992020701448994](https://doi.org/10.1080/14992020701448994).
- Horowitz-Kraus, T., Grainger, M., DiFrancesco, M., Vannest, J., Holland, S.K., 2015. Right is not always wrong: DTI and fMRI evidence for the reliance of reading comprehension on language-comprehension networks in the right hemisphere. *Brain Imaging Behav.* 9 (1), 19–31. doi:[10.1007/s11682-014-9341-9](https://doi.org/10.1007/s11682-014-9341-9).
- Humphries, C., Binder, J.R., Medler, D.A., Liebenthal, E., 2007. Time course of semantic processes during sentence comprehension: an fMRI study. *NeuroImage* 36 (3), 924–932. doi:[10.1016/j.neuroimage.2007.03.059](https://doi.org/10.1016/j.neuroimage.2007.03.059).
- Huth, A.G., de Heer, W.A., Griffiths, T.L., Theunissen, F.E., Gallant, J.L., 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* 532 (7600), 453–458. doi:[10.1038/nature17637](https://doi.org/10.1038/nature17637).
- Huth, A.G., Nishimoto, S., Vu, A.T., Gallant, J.L., 2012. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron* 76 (6), 1210–1224. doi:[10.1016/j.neuron.2012.10.014](https://doi.org/10.1016/j.neuron.2012.10.014).
- Hwang, K., Hallquist, M.N., Luna, B., 2013. The development of hub architecture in the human functional brain network. *Cereb. Cortex* 23 (10), 2380–2393. doi:[10.1093/cercor/crs227](https://doi.org/10.1093/cercor/crs227).
- Ioannidis, J.P.A., 2005. Why most published research findings are false. *PLoS Med.* 2 (8), e124. doi:[10.1371/journal.pmed.0020124](https://doi.org/10.1371/journal.pmed.0020124).
- Jackson, R.L., 2021. The neural correlates of semantic control revisited. *NeuroImage* 224, 117444. doi:[10.1016/j.neuroimage.2020.117444](https://doi.org/10.1016/j.neuroimage.2020.117444).
- Kersey, A.J., Clark, T.S., Lussier, C.A., Mahon, B.Z., Cantlon, J.F., 2016. Development of tool representations in the dorsal and ventral visual object processing pathways. *Cereb. Cortex* 26 (7), 3135–3145. doi:[10.1093/cercor/bhv140](https://doi.org/10.1093/cercor/bhv140).

- Krishnan, S., Asaridou, S.S., Cler, G.J., Smith, H.J., Willis, H.E., Healy, M.P., Thompson, P.A., Bishop, D.V.M., Watkins, K.E., 2021. Functional organisation for verb generation in children with developmental language disorder. *NeuroImage* 226, 117599. doi:[10.1016/j.neuroimage.2020.117599](https://doi.org/10.1016/j.neuroimage.2020.117599).
- Krishnan, S., Leech, R., Mercure, E., Lloyd-Fox, S., Dick, F., 2015. Convergent and divergent fMRI responses in children and adults to increasing language production demands. *Cereb. Cortex* 25 (10), 3261–3277. doi:[10.1093/cercor/bhu120](https://doi.org/10.1093/cercor/bhu120).
- Laird, A.R., Fox, P.M., Price, C.J., Glahn, D.C., Uecker, A.M., Lancaster, J.L., Turkeltaub, P.E., Kochunov, P., Fox, P.T., 2005. ALE meta-analysis: controlling the false discovery rate and performing statistical contrasts. *Hum. Brain Mapp.* 25 (1), 155–164. doi:[10.1002/hbm.20136](https://doi.org/10.1002/hbm.20136).
- Ralph, Lambon, A., M., Jefferies, E., Patterson, K., Rogers, T.T., 2017. The neural and computational bases of semantic cognition. *Nat. Rev. Neurosci.* 18 (1), 42–55. doi:[10.1038/nrn.2016.150](https://doi.org/10.1038/nrn.2016.150).
- Lancaster, J.L., Tordesillas-Gutiérrez, D., Martínez, M., Salinas, F., Evans, A., Zilles, K., Mazziotta, J.C., Fox, P.T., 2007. Bias between MNI and Talairach coordinates analyzed using the ICBM-152 brain template. *Hum. Brain Mapp.* 28 (11), 1194–1205. doi:[10.1002/hbm.20345](https://doi.org/10.1002/hbm.20345).
- Leach, J.L., Holland, S.K., 2010. Functional MRI in children: clinical and research applications. *Pediatr. Radiol.* 40 (1), 31–49. doi:[10.1007/s00247-009-1452-x](https://doi.org/10.1007/s00247-009-1452-x). Scopus.
- Lee, S.-H., Booth, J.R., Chen, S.-Y., Chou, T.-L., 2011. Developmental changes in the inferior frontal cortex for selecting semantic representations. *Dev. Cogn. Neurosci.* 1 (3), 338–350. doi:[10.1016/j.dcn.2011.01.005](https://doi.org/10.1016/j.dcn.2011.01.005).
- Lee, S.-H., Booth, J.R., Chou, T.-L., 2016. Temporo-parietal connectivity uniquely predicts reading change from childhood to adolescence. *NeuroImage* 142, 126–134. doi:[10.1016/j.neuroimage.2016.06.055](https://doi.org/10.1016/j.neuroimage.2016.06.055).
- Libertus, M.E., Brannon, E.M., Pelpfrey, K.A., 2009. Developmental changes in category-specific brain responses to numbers and letters in a working memory task. *NeuroImage* 44 (4), 1404–1414. doi:[10.1016/j.neuroimage.2008.10.027](https://doi.org/10.1016/j.neuroimage.2008.10.027).
- Liebig, J., Froehlich, E., Morawetz, C., Braun, M., Jacobs, A.M., Heekeren, H.R., Ziegler, J.C., 2017. Neurofunctional dissection of the reading system in children. *Dev. Cogn. Neurosci.* 27, 45–57. doi:[10.1016/j.dcn.2017.07.002](https://doi.org/10.1016/j.dcn.2017.07.002).
- Liuzzi, A.G., Aglinskas, A., Fairhall, S.L., 2020. General and feature-based semantic representations in the semantic network. *Sci. Rep.* 10 (1), 8931. doi:[10.1038/s41598-020-65906-0](https://doi.org/10.1038/s41598-020-65906-0).
- Martin, A., 2016. GRAPES—grounding representations in action, perception, and emotion systems how object properties and categories are represented in the human brain. *Psychon. Bull. Rev.* 23 (4), 979–990. doi:[10.3758/s13423-015-0842-3](https://doi.org/10.3758/s13423-015-0842-3).
- Martin, A., Schurz, M., Kronbichler, M., Richlan, F., 2015. Reading in the brain of children and adults: a meta-analysis of 40 functional magnetic resonance imaging studies. *Hum. Brain Mapp.* 36 (5), 1963–1981. doi:[10.1002/hbm.22749](https://doi.org/10.1002/hbm.22749).
- Mather, A., Schultz, D., Wang, Y., 2020. Neural bases of phonological and semantic processing in early childhood. *Brain Connect.* 10 (5), 212–223. doi:[10.1089/brain.2019.0728](https://doi.org/10.1089/brain.2019.0728).
- Meyler, A., Keller, T.A., Cherkassky, V.L., Gabrieli, J.D.E., Just, M.A., 2008. Modifying the brain activation of poor readers during sentence comprehension with extended remedial instruction: a longitudinal study of neuroplasticity. *Neuropsychologia* 46 (10), 2580–2592. doi:[10.1016/j.neuropsychologia.2008.03.012](https://doi.org/10.1016/j.neuropsychologia.2008.03.012).
- Monzalvo, K., Fluss, J., Billard, C., Dehaene, S., Dehaene-Lambertz, G., 2012. Cortical networks for vision and language in dyslexic and normal children of variable socio-economic status. *NeuroImage* 61 (1), 258–274. doi:[10.1016/j.neuroimage.2012.02.035](https://doi.org/10.1016/j.neuroimage.2012.02.035).
- Moore-Parks, E.N., Burns, E.L., Bazzill, R., Levy, S., Posada, V., Müller, R.-A., 2010. An fMRI study of sentence-embedded lexical-semantic decision in children and adults. *Brain Lang.* 114 (2), 90–100. doi:[10.1016/j.bandl.2010.03.009](https://doi.org/10.1016/j.bandl.2010.03.009).
- Müller, V.I., Cieslik, E.C., Laird, A.R., Fox, P.T., Radua, J., Mataix-Cols, D., Tench, C.R., Yarkoni, T., Nichols, T.E., Turkeltaub, P.E., Wager, T.D., Eickhoff, S.B., 2018. Ten simple rules for neuroimaging meta-analysis. *Neurosci. Biobehav. Rev.* 84, 151–161. doi:[10.1016/j.neubiorev.2017.11.012](https://doi.org/10.1016/j.neubiorev.2017.11.012).
- Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.-B., 2005. Valid conjunction inference with the minimum statistic. *NeuroImage* 25 (3), 653–660. doi:[10.1016/j.neuroimage.2004.12.005](https://doi.org/10.1016/j.neuroimage.2004.12.005).
- Noonan, K.A., Jefferies, E., Visser, M., Ralph, Lambon, M., A., 2013. Going beyond inferior prefrontal involvement in semantic control: evidence for the additional contribution of dorsal angular gyrus and posterior middle temporal cortex. *J. Cogn. Neurosci.* 25 (11), 1824–1850. doi:[10.1162/jocn_a_00442](https://doi.org/10.1162/jocn_a_00442).
- Notter, M.P., Gale, D., Herholz, P., Markele, R., Notter-Bielser, M.-L., Whitaker, K., 2019. AtlasReader: a Python package to generate coordinate tables, region labels, and informative figures from statistical MRI images. *J. Open Source Softw.* 4 (34), 1257. doi:[10.21105/joss.01257](https://doi.org/10.21105/joss.01257).
- Núñez, S.C., Dapretto, M., Katzir, T., Starr, A., Bramen, J., Kan, E., Bookheimer, S., Sowell, E.R., 2011. fMRI of syntactic processing in typically developing children: structural correlates in the inferior frontal gyrus. *Dev. Cogn. Neurosci.* 1 (3), 313–323. doi:[10.1016/j.dcn.2011.02.004](https://doi.org/10.1016/j.dcn.2011.02.004).
- Okamoto, Y., Kosaka, H., Kitada, R., Seki, A., Tanabe, H.C., Hayashi, M.J., Kochiyama, T., Saito, D.N., Yanaka, H.T., Munesue, T., Ishitobi, M., Omori, M., Wada, Y., Okazawa, H., Koeda, T., Sadato, N., 2017. Age-dependent atypicalities in body- and face-sensitive activation of the EBA and FFA in individuals with ASD. *Neurosci. Res.* 119, 38–52. doi:[10.1016/j.neures.2017.02.001](https://doi.org/10.1016/j.neures.2017.02.001).
- O'Shaughnessy, E.S., Berl, M.M., Moore, E.N., Gaillard, W.D., 2008. Pediatric functional magnetic resonance imaging (fMRI): issues and applications. *J. Child Neurol.* 23 (7), 791–801. doi:[10.1177/0883073807313047](https://doi.org/10.1177/0883073807313047).
- Passarotti, A.M., Paul, B.M., Bussière, J.R., Buxton, R.B., Wong, E.C., Stiles, J., 2003. The development of face and location processing: an fMRI study. *Dev. Sci.* 6 (1), 100–117. doi:[10.1111/1467-7687.00259](https://doi.org/10.1111/1467-7687.00259).
- Patterson, K., Nestor, P.J., Rogers, T.T., 2007. Where do you know what you know? the representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 8 (12), 976–987. doi:[10.1038/nrn2277](https://doi.org/10.1038/nrn2277).
- Poline, J.-B., Breeze, J.L., Ghosh, S., Gorgolewski, K., Halchenko, Y.O., Hanke, M., Haselgrave, C., Helmer, K.G., Keator, D.B., Marcus, D.S., Poldrack, R.A., Schwartz, Y., Ashburner, J., Kennedy, D.N., 2012. Data sharing in neuroimaging research. *Front. Neuroinform.* 6, 9. doi:[10.3389/fninf.2012.00009](https://doi.org/10.3389/fninf.2012.00009).
- Pulvermüller, F., Fadiga, L., 2010. Active perception: sensorimotor circuits as a cortical basis for language. *Nat. Rev. Neurosci.* 11 (5), 351–360. doi:[10.1038/nrn2811](https://doi.org/10.1038/nrn2811).
- Pulvermüller, F., Kherif, F., Hauk, O., Mohr, B., Nimmo-Smith, I., 2009. Distributed cell assemblies for general lexical and category-specific semantic processing as revealed by fMRI cluster analysis. *Hum. Brain Mapp.* 30 (12), 3837–3850. doi:[10.1002/hbm.20811](https://doi.org/10.1002/hbm.20811).
- Rodd, J.M., Vitello, S., Woollams, A.M., Adank, P., 2015. Localising semantic and syntactic processing in spoken and written language comprehension: an activation likelihood estimation meta-analysis. *Brain Lang.* 141, 89–102. doi:[10.1016/j.bandl.2014.11.012](https://doi.org/10.1016/j.bandl.2014.11.012).
- Rolls, E.T., Joliot, M., Tzourio-Mazoyer, N., 2015. Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas. *NeuroImage* 122, 1–5. doi:[10.1016/j.neuroimage.2015.07.075](https://doi.org/10.1016/j.neuroimage.2015.07.075).
- Rosenthal, R., 1979. The file drawer problem and tolerance for null results. *Psychol. Bull.* 86 (3), 638–641. doi:[10.1037/0033-2909.86.3.638](https://doi.org/10.1037/0033-2909.86.3.638).
- Sachs, B.C., Gaillard, W.D., 2003. Organization of language networks in children: functional magnetic resonance imaging studies. *Curr. Neurol. Neurosci. Rep.* 3 (2), 157–162. doi:[10.1007/s11910-003-0068-z](https://doi.org/10.1007/s11910-003-0068-z).
- Salimi-Khorshidi, G., Smith, S.M., Keltner, J.R., Wager, T.D., Nichols, T.E., 2009. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *NeuroImage* 45 (3), 810–823. doi:[10.1016/j.neuroimage.2008.12.039](https://doi.org/10.1016/j.neuroimage.2008.12.039).
- Salo, T., Yarkoni, T., Bottemhorn, K., Nichols, T., Gorgolewski, K., Riedel, M., Kent, J., Glerean, E., Bilgel, M., Wright, J., Reeder, P., Nielson, D., Yanes, J., Pérez, A., Sutherland, M., Laird, A., 2020. NiMARE: a neuroimaging meta-analysis research environment [Poster presentation]. OHBM Annual Meeting July 23–July 3 doi:[10.5281/zenodo.4562954](https://doi.org/10.5281/zenodo.4562954).
- Samartisidis, P., Montagna, S., Laird, A.R., Fox, P.T., Johnson, T.D., Nichols, T.E., 2020. Estimating the prevalence of missing experiments in a neuroimaging meta-analysis. *Res. Synth. Methods* 11 (6), 866–883. doi:[10.1002/jrsm.1448](https://doi.org/10.1002/jrsm.1448).
- Samartisidis, P., Montagna, S., Nichols, T.E., Johnson, T.D., 2017. The coordinate-based meta-analysis of neuroimaging data. *Stat. Sci.* 32 (4), 580–599. doi:[10.1214/17-STS624](https://doi.org/10.1214/17-STS624).
- Schafer, R.J., Lacadie, C., Vohr, B., Kesler, S.R., Katz, K.H., Schneider, K.C., Pugh, K.R., Makuch, R.W., Reiss, A.L., Constable, R.T., Ment, L.R., 2009. Alterations in functional connectivity for language in prematurely born adolescents. *Brain* 132 (3), 661–670. doi:[10.1093/brain/awn353](https://doi.org/10.1093/brain/awn353).
- Scherf, K.S., Behrmann, M., Humphreys, K., Luna, B., 2007. Visual category-selectivity for faces, places and objects emerges along different developmental trajectories. *Dev. Sci.* 10 (4), F15–F30. doi:[10.1111/j.1467-7687.2007.00595.x](https://doi.org/10.1111/j.1467-7687.2007.00595.x).
- Scherf, K.S., Luna, B., Minshew, N., Behrmann, M., 2010. Location, location, location: alterations in the functional topography of face- but not object- or place-related cortex in adolescents with autism. *Front. Hum. Neurosci.* 4, 26. doi:[10.3389/fnhum.2010.00026](https://doi.org/10.3389/fnhum.2010.00026).
- Schlaggar, B.L., McCandllis, B.D., 2007. Development of neural systems for reading. *Annu. Rev. Neurosci.* 30 (1), 475–503. doi:[10.1146/annurev.neuro.28.061604.135645](https://doi.org/10.1146/annurev.neuro.28.061604.135645).
- Siok, W.T., Perfetti, C.A., Jin, Z., Tan, L.H., 2004. Biological abnormality of impaired reading is constrained by culture. *Nature* 431 (7004), 71–76. doi:[10.1038/nature02865](https://doi.org/10.1038/nature02865).
- Skeide, M.A., Brauer, J., Friederici, A.D., 2014. Syntax gradually segregates from semantics in the developing brain. *NeuroImage* 100, 106–111. doi:[10.1016/j.neuroimage.2014.05.080](https://doi.org/10.1016/j.neuroimage.2014.05.080).
- Skeide, M.A., Friederici, A.D., 2016. The ontogeny of the cortical language network. *Nat. Rev. Neurosci.* 17 (5), 323–332. doi:[10.1038/nrn.2016.23](https://doi.org/10.1038/nrn.2016.23). Scopus.
- Stevens, M.C., Skudlarski, P., Pearlson, G.D., Calhoun, V.D., 2009. Age-related cognitive gains are mediated by the effects of white matter development on brain network integration. *NeuroImage* 48 (4), 738–746. doi:[10.1016/j.neuroimage.2009.06.065](https://doi.org/10.1016/j.neuroimage.2009.06.065).
- Szaflarski, J.P., Schmitzhorst, V.J., Altaye, M., Byars, A.W., Ret, J., Plante, E., Holland, S.K., 2006. A longitudinal functional magnetic resonance imaging study of language development in children 5 to 11 years old. *Ann. Neurol.* 59 (5), 796–807. doi:[10.1002/ana.20817](https://doi.org/10.1002/ana.20817).
- Thirion, B., Pinel, P., Mériaux, S., Roche, A., Dehaene, S., Poline, J.-B., 2007. Analysis of a large fMRI cohort: statistical and methodological issues for group analyses. *NeuroImage* 35 (1), 105–120. doi:[10.1016/j.neuroimage.2006.11.054](https://doi.org/10.1016/j.neuroimage.2006.11.054).
- Travis, K.E., Leonard, M.K., Brown, T.T., Hagler Jr., D.J., Curran, M., Dale, A.M., Elman, J.L., Halgren, E., 2011. Spatiotemporal neural dynamics of word understanding in 12-to 18-month-old infants. *Cereb. Cortex* 21 (8), 1832–1839. doi:[10.1093/cercor/cbh259](https://doi.org/10.1093/cercor/cbh259).
- Turkeltaub, P.E., Eden, G.F., Jones, K.M., Zeffiro, T.A., 2002. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *NeuroImage* 16 (3), 765–780. doi:[10.1006/nimg.2002.1131](https://doi.org/10.1006/nimg.2002.1131).
- Turkeltaub, P.E., Eickhoff, S.B., Laird, A.R., Fox, M., Wiener, M., Fox, P., 2012. Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Hum. Brain Mapp.* 33 (1), 1–13. doi:[10.1002/hbm.21186](https://doi.org/10.1002/hbm.21186).
- Tyler, L.K., Stamatakis, E.A., Dick, E., Bright, P., Fletcher, P., Moss, H., 2003. Objects and their actions: evidence for a neurally distributed semantic system. *NeuroImage* 18 (2), 542–557. doi:[10.1016/S1053-8119\(02\)00047-2](https://doi.org/10.1016/S1053-8119(02)00047-2).
- Van Rossum, G., Drake, F.L., 2009. Python 3 Reference Manual. CreateSpace.

- Vannest, J., Szafarski, J.P., Eaton, K.P., Henkel, D.M., Morita, D., Glauser, T.A., Byars, A.W., Patel, K., Holland, S.K., 2012. Functional magnetic resonance imaging reveals changes in language localization in children with benign childhood epilepsy with centrotemporal spikes. *J. Child Neurol.* 28 (4), 435–445. doi:[10.1177/0883073812447682](https://doi.org/10.1177/0883073812447682).
- Vigneau, M., Beaucousin, V., Hervé, P.Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., Tzourio-Mazoyer, N., 2006. Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *NeuroImage* 30 (4), 1414–1432. doi:[10.1016/j.neuroimage.2005.11.002](https://doi.org/10.1016/j.neuroimage.2005.11.002).
- Visser, M., Jefferies, E., Lambon Ralph, M.A., 2010. Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *J. Cogn. Neurosci.* 22 (6), 1083–1094. doi:[10.1162/jocn.2009.21309](https://doi.org/10.1162/jocn.2009.21309).
- Weiss-Croft, L.J., Baldeweg, T., 2015. Maturation of language networks in children: a systematic review of 22 years of functional MRI. *NeuroImage* 123, 269–281. doi:[10.1016/j.neuroimage.2015.07.046](https://doi.org/10.1016/j.neuroimage.2015.07.046).
- Wong, C.-H., Gau, S.S.-F., Chou, T.-L., 2019. Neural correlates of association strength and categorical relatedness in youths with autism spectrum disorder. *Autism Res.* 12 (10), 1484–1494. doi:[10.1002/aur.2184](https://doi.org/10.1002/aur.2184).
- Wu, C.-Y., Ho, M.H.R., Chen, S.H.A., 2012. A meta-analysis of fMRI studies on Chinese orthographic, phonological, and semantic processing. *NeuroImage* 63 (1), 381–391. doi:[10.1016/j.neuroimage.2012.06.047](https://doi.org/10.1016/j.neuroimage.2012.06.047).
- Wu, C.-Y., Vissiennon, K., Friederici, A.D., Brauer, J., 2016. Preschoolers' brains rely on semantic cues prior to the mastery of syntax during sentence comprehension. *NeuroImage* 126, 256–266. doi:[10.1016/j.neuroimage.2015.10.036](https://doi.org/10.1016/j.neuroimage.2015.10.036).
- Xue, G., Dong, Q., Jin, Z., Zhang, L., Wang, Y., 2004. An fMRI study with semantic access in low proficiency second language learners. *NeuroReport* 15 (5), 791–796. doi:[10.1097/00001756-200404090-00010](https://doi.org/10.1097/00001756-200404090-00010).
- Yarkoni, T., 2021. The generalizability crisis. *Behav. Brain Sci.* 1–37. doi:[10.1017/S0140525X20001685](https://doi.org/10.1017/S0140525X20001685).

Tracking the neural correlates of learning to read with dense-sampling fMRI

Alexander Enge^{1,2} & Michael A. Skeide¹

¹ Max Planck Institute for Human Cognitive and Brain Sciences

² Humboldt-Universität zu Berlin

Learning to read is a developmental milestone of lifelong importance. While many studies have demonstrated the neural correlates of reading in proficient adult readers, neuroimaging studies of learning to read during childhood are relatively scarce, and have not managed to disentangle effects specific to reading instruction from effects of general cognitive development and schooling. Furthermore, previous reading research has almost exclusively focused on participants and writing systems in the Global North. To overcome these limitations, we conducted a dense-sampling longitudinal fMRI study with children from a socio-economically disadvantaged background ($N = 15$; age = 5–9 years) who participated in approximately 16 months of reading instruction in their native language (Hindi/Devanagari). During each of up to six fMRI scanning sessions, spaced at intervals of approximately 2–3 months, children were presented with spoken and written word-like stimuli, including low-level sensory control stimuli (noise-vocoded speech and false fonts), pseudowords, and words. We estimated longitudinal change in BOLD activity amplitude and BOLD activity patterns using linear mixed-effects models. We found few changes (both linear and non-linear) in BOLD activity amplitudes for different contrasts and brain areas, including an inverted “u”-shaped pattern of activity increase and decrease for written words compared to visual low-level control stimuli. However, none of these changes occurred in anatomically plausible areas based on previous neuroimaging studies. We also found an increase in audio-visual multi-voxel pattern similarity for written words as compared low-level sensory controls in the left ventral occipito-temporal (vOT) cortex. There were no reliable changes in word- or pseudoword-related multi-voxel pattern stability. This lack of significant findings may be due to limitations in our study design, including a very small sample size and a reading intervention that might have been too short or too weak. Nevertheless, we hope that our rationale for this study as well as our whole-brain longitudinal analysis pipeline will inspire future cognitive-neuroscientific research.

Keywords: reading, literacy, learning, children, fMRI, neuroimaging

Introduction

Learning to read is a developmental milestone of lifelong importance. Reading and writing enable us to exchange messages across time and space, to remember things we would otherwise forget, to educate ourselves and others, and to enjoy fictional stories. Without it, we would likely be missing out on many of humankind’s greatest achievements, including science and technology. It therefore seems fair to say that the process of learning to read, typically taking place during late kindergarten and early primary school education as well as through the help of parents and other caretakers, is a crucial step in children’s education and cognitive development.

Children typically pick up knowledge about their first letters informally, for example by being shown how their own name is written or by memorizing the logo of their favorite TV series or cereal brand. However, at this first, *logographic* stage of reading development (Ehri, 2005; Frith, 1986), children are typically not yet aware of the systematic correspondence between individual letters (graphemes) and speech sounds (phonemes), and therefore unable to pronounce or understand novel words. In this stage, written words are probably perceived similar to other types of visual objects, and novel words that have never been encountered before do not elicit any phonological or semantic representation, as they would for a trained reader.

Through structured training in kindergarten, primary school, or at home, children advance to the second, *alphabetic* stage of reading development,¹ where they are taught that individual letters are associated with individual speech sounds. This allows children to process the

The analysis code for this study is openly available at <https://github.com/SkeideLab/SLANG>.

We have no conflict of interest to disclose.

Correspondence concerning this article should be addressed to Alexander Enge, Stephanstraße 1a, 04103 Leipzig, Germany. E-mail: alex_enge@web.de

¹Note, however, that as with any model of developmental “stages” or “phases,” there typically is no hard and clear-cut boundary between one stage and the next, and there is typically large interindividual variation in the onsets and durations of the stages.

individual letters of a written word sequentially, convert them to their corresponding speech sounds, and combine those speech sounds to vocalize the entire word. Since most children will have become proficient in spoken language comprehension long before learning to read (typically with < 3 years of age; Skeide & Friederici, 2016), they will also be able to access the meaning of the written word based on its assembled spoken word form. For this way of reading to be successful, children need to possess the understanding and skill to break down words into their constituent phonemes, typically referred to as “phonological awareness.” It is therefore unsurprising that phonological skills like first-letter naming, rhyming, and verbal short-term memory are among the best predictors of individual differences in future reading ability (e.g., Melby-Lervåg et al., 2012). However, this alphabetic form of reading remains somewhat slow and error-prone due to the sequential conversion of letters into speech sounds.

Only in the third and final stage of reading development, the *orthographic* stage, children become able to read words as a whole and to access their meaning directly from the visual word form, without having to take the detour of phonological decoding. Presumably, it is the emergence of this shortcut that substantially reduces word reading times after approximately the first year of reading instruction (e.g., Hasenäcker et al., 2017) and which makes adult reading so efficient and seemingly effortless.

Neural correlates of reading

Reading as a cognitive function has to be implemented on a neuroanatomical and functional level in the brain. To probe which brain areas contribute to different aspects reading in proficient readers, cognitive-neuroscientific methods such as electroencephalography (EEG) and magnetic resonance imaging (MRI) can be used to measure brain activity while participants perform reading tasks. Due to its comparatively high spatial resolution (on the order of a few millimeters), functional magnetic resonance imaging (fMRI), which measures changes in blood oxygen level as a proxy of local neuronal activity, can be used to isolate which brain areas are more active during reading than during other control tasks.

One well-replicated finding using this methodology is that processing written words engages a relatively circumscribed region on the left ventral occipito-temporal (vOT) cortex. This region, typically referred to as the visual word form area (VWFA; e.g., Cohen et al., 2000; Dehaene et al., 2002; Dehaene & Cohen, 2011; McCandliss et al., 2003), is thought to be a purely visual and pre-lexical area that identifies written words based on its lower-level visual shapes. There is converging evidence that anomalies in, damage to, and stimulation of the VWFA can cause reading difficulties (Brem et al., 2020; Hillis et al., 2005; Hirshorn et al., 2016; but see also Price & Devlin, 2003). Furthermore, the VWFA is often described as a prime example of the neuronal recycling hypothesis (Dehaene & Cohen, 2007), according to which “modern” cognitive functions such as reading,² for which

not enough evolutionary time has passed to develop dedicated brain circuitry, repurpose brain areas with similar but evolutionarily older functions. In the case of reading, the VWFA may reuse regions that had previously been specialized for recognizing other complex visual object categories such as faces, limbs, or tools (Dehaene et al., 2015; Kubota et al., 2023; Nordt et al., 2021; but see also Coltheart, 2014).

The brain areas involved in reading beyond the early stages of visual word form recognition are less well understood, presumably owing to large differences in task design between studies. A meta-analysis of fMRI study (Murphy et al., 2019) found that beyond the VWFA, single word reading reliable engages a left-lateralized set of regions including the left inferior frontal and left superior and middle temporal gyri, all of which are known to be involved in phonological and semantic processing of spoken language. However, it remains an open question which brain areas serve as the interface between visual processing (word form recognition in the VWFA) and language processing (phonological and semantic processing in the left perisylvian language network) during reading. One candidate for the visual–phonological interface, that is, linking written letters to speech sounds, is the left posterior superior temporal gyrus (pSTS). This area has been shown to integrate auditory and visual information when both are presented concurrently, e.g., during audio-visual letter perception or lip reading tasks (e.g., Blau et al., 2010; Calvert et al., 1997; van Atteveldt et al., 2004; Wilson et al., 2018).³ One candidate for the visual–semantic interface is the left middle fusiform cortex (lmFFC), which lies anterior to the high-level visual object recognition areas (including the VWFA) on the ventral surface of the occipital and temporal lobes. Using depth electrodes for recording and stimulation in epileptic patients,⁴ it has been shown that this region is active for lexical retrieval from both auditory input (spoken words) and visual input (written words and object images; e.g., Forseth et al., 2018; Woolnough et al., 2020, 2022).

Neural correlates of learning to read

Compared to hundreds if not thousands of studies in proficient adult readers, there has been a lot less research on how the neural correlates of reading develop in beginning readers. There are multiple reasons for this: (a) It is much harder to recruit children (and their families), as compared to undergraduate students, who depend on obtaining course credit or monetary compensation; (b) Children in the relevant age range (late kindergarten to early primary school; typically ~5–8 years of age) have a shorter attention span and show more in-scanner head

²The first known scripts appeared approximately 3000–5000 B.C.E (e.g., Houston, 2004).

³Note that if the left pSTS does indeed play a role in reading, this could be viewed as another case of neuronal recycling, since its “new” skill of linking written letters to speech sounds may stem from its “original” skill of linking lip movements to speech sounds.

⁴Unfortunately, the depth of this region and the associated signal dropout makes it hard to pick up BOLD activity changes with fMRI (Embleton et al., 2010; Liu, 2016).

movement, leading to fewer usable scans, lower data quality in usable scans, and fewer data points per scan; and (c) Accurately tracking the development of brain structure and function in individual children requires a longitudinal study design, which is very costly and demanding for study participants, typically leading to small sample sizes. Nevertheless, a few studies exist that have managed to obtain multiple longitudinal scans of children's brain activity as they are learning to read.

Dehaene-Lambertz et al. (2018) scanned ten 6-year-old children longitudinally throughout the first year of schooling (6–7 scans per child) and presented them with visual objects from different object categories, including written words. Behaviorally, they found that reading performance (knowledge of grapheme–phoneme relations and number of words read per minutes in a standardized reading test) increased sharply during the first months of schooling. In the fMRI, they found that selectivity for words in the VWFA was not present at the beginning of the study but quickly emerged within less than 6 months in most of the children. Interestingly, the activation strength in the VWFA and other word-selective areas appeared to follow a curvilinear inverted “u”-shaped pattern, with a quick rise in BOLD activation strength in the first half of the study followed by a slight decline in the second half. Such a pattern would be predicted by the *expansion and renormalization* model of brain plasticity (Wenger et al., 2017), according to which a novel skill initially requires more resources (in terms of number of voxels or BOLD activation amplitude) but later on becomes more efficient and automatized. Regarding the neuronal recycling hypothesis, Dehaene-Lambertz et al. (2018) found that the voxels that would later form the VWFA in individual children were weakly tuned to a different object category at the beginning of the study, namely tools. However, a few shortcomings of this study need to be noted: (a) The sample size (both in terms of number of children and number of scans per child) was very small, therefore leading to low statistical power to detect all but very large effects (see also Button et al., 2013; Ioannidis, 2005; Szucs & Ioannidis, 2017); (b) Their analysis of variance (ANOVA) model did not take into account the longitudinal nature of the data, with repeated measures of the same participants likely being positively correlated with one another; (c) It is unclear if the changes observed in this study were caused by reading instruction in particular or by other aspects of schooling or general cognitive development, since there was no non-reading control group (for obvious practical and ethical reasons); and (d) The study design captured only the very first stage of reading, namely visual word form recognition, while it remains an open question how visual word forms get linked to other aspects of (spoken) language comprehension, namely speech sounds (phonology) and word meaning (semantics).

Cultural biases in reading research

Most of psychological and cognitive-neuroscientific research is carried out in countries of the Global North, especially in Western Europe and Northern America. Within these countries, study participants are not sam-

pled at random, but typically come from economically and educationally privileged social backgrounds (e.g., psychology students). This restriction of study samples can limit the generalizability of research findings, as even basic and presumably “universal” effects such as some perceptual illusions do not necessarily replicate in participants from different cultural and socio-economic backgrounds (Blasi et al., 2022; Henrich et al., 2010).

In reading research, this problem is potentiated by the fact that different cultures developed—sometimes radically—different writing systems. Due to the concentration of research funds and technology in the Global North, research on reading, its development, and its neural correlates has almost exclusively been carried out in languages with *alphabetic* writing systems such as English, German, or French. In these writing systems, individual units of written language (graphemes) correspond to very small units of spoken language (phonemes).⁵ On the other end of the spectrum are *logographic* writing systems (such Chinese), in which individual graphemes correspond to large units of spoken languages, namely entire words or concepts (morphemes). In between these two extremes sit *syllabic* writing systems, in which individual graphemes correspond to intermediate units of spoken languages, namely sublexical consonant–vowel combinations (syllables). A special case are *alphasyllabic* languages (also “abugidas”; such as the Hindi writing system Devanagari), in which individual graphemes correspond to consonants with added diacritical marks above, below, or next to the letter for vowels.

There is an abundance of research findings on different aspects of reading and its neural correlates in alphabetic languages,⁶ but only very few studies that tested logographic writing systems (typically Chinese) and next to none that tested syllabic or alphasyllabic writing systems. These biases clearly limit the scope of empirical findings and current theories on reading and its development to a small set of regions and languages (Frost, 2012; Share, 2008, 2014, 2021).

The present study

Our goal here was to capture the developmental changes in brain activity related to written word processing as children are learning to read. To this end, we conducted a longitudinal fMRI study with children at 5–9 years of age who received explicit reading instruction for approximately 1.5 years. Children were scanned at relatively short intervals of approximately 2–3 months for a total of up to 6 scanning sessions per child. At each scanning session, children were presented with spoken and written words, pseudowords, and low-level sensory control stimuli. This allowed us to capture intra-individual changes

⁵Though the tightness/reliability of this correspondence differs between relatively shallow (transparent) orthographies such as Spanish, Italian, or German, and relatively deep (opaque) orthographies such as English and French.

⁶Unfortunately, most studies implicitly claim generalizability/universality of their findings by not explicitly mentioning the writing system in the title, abstract, or conclusion of the paper.

in BOLD activity (as a proxy for local neuronal activity) in response to stimuli that differed in their sensory, phonological, and semantic content. To overcome the bias towards study participants from the Global North and alphabetic writing systems, we tested children from socio-economically disadvantaged backgrounds in India who received reading instruction in their native language (Hindi) and alphasyllabic writing system (Devanagari). Our hypotheses were the following:

- We expected that BOLD activity in response to written (pseudo-)words would increase as children are learning to read, either in a linear or quadratic (inverted “u”-shaped) pattern. This was based on previous findings (e.g., Dehaene-Lambertz et al., 2018) demonstrating the quick emergence of word selectivity in higher-level visual cortex.
- We expected that in audio-visual integration areas (especially the pSTS and vOT cortex), multi-voxel BOLD activity patterns would become more similar between written and spoken (pseudo-)words as children are learning to read. This was based on the intuition that only over the course of the study, children would become able to access the phonological and semantic content of written words.
- We expected that multi-voxel BOLD activity patterns for written (pseudo-)words would become more “stable,” i.e., show less stimulus-to-stimulus variation as children are learning to read. This was based on the intuition that activity in reading-related areas should become more finely tuned and less noisy towards written words as children become able to decode and comprehend them.

Methods

Participants

We initially recruited a total of 32 children from a small village in the region of Uttar Pradesh, India, to participate in the reading intervention. In a neighboring village, we recruited an additional 25 children to participate in a mathematics intervention, serving as an active control group. However, data from these children was not analyzed for the purpose of the present manuscript. The two villages were selected in cooperation with a local non-governmental organization, based on there being little to no access to primary school education due to geographic and socio-demographic constraints. All recruited children had to fulfil the following inclusion criteria: (1) age between 5 and 9 years at the beginning of the study, (2) not being able to decode letters and/or read words, (3) not attending school, (4) not fulfilling any contraindication for MRI scanning (e.g., no relevant implants or medication), (5) no hearing and/or vision impairment, (6) no language impairments, and (7) no attention deficits. From the 32 children initially recruited for the reading intervention group, 17 children were excluded because they did not participate in at least two scanning sessions or because they did not meet our cutoff criterion for head motion (framewise displacement greater than 0.5 mm in less than 25% of fMRI volumes) in at least two scanning sessions. Therefore, the final sample size was $N = 15$ children. Of

those, 9 identified as female and 6 identified as male. The mean age at the beginning of the study was 7.13 years ($SD = 1.25$ years, min = 5 years, max = 9 years).

Study design

The children in this longitudinal study completed regular MRI scanning sessions while participating in a structured reading program (or mathematics program for the control group, not included in this manuscript). The first scanning session took place at the beginning of the program and the next scanning sessions (mean = 5.07 sessions per child, $SD = 1.53$ sessions, min = 2 sessions, max = 6 sessions) were spaced at intervals of approximately 2–3 months (mean = 79.9 days between sessions, $SD = 35.8$ days, min = 20 days, max = 182 days; see Figure 1). Each scanning session consisted of one localizer scan, one functional MRI scan (see experimental design and scanning parameters below), one structural MRI scan (see scanning parameters below), and a series of standardized behavioral tests outside of the scanner (see behavioral testing below).

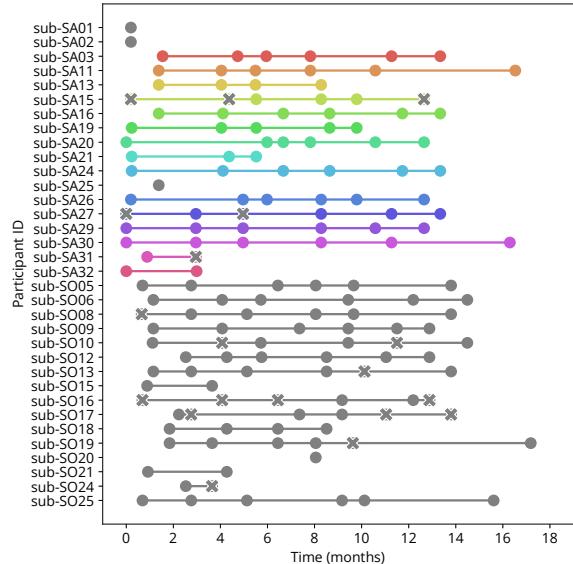


Figure 1. Longitudinal data acquisition schedule. Colored dots represent data acquisition points (MRI and behavioral tests) for children from the reading intervention group (participants “sub-SA01” to “sub-SA32”). Gray dots represent data acquisition points (MRI and behavioral tests) from the mathematics control group (participants “sub-SO01” to “sub-SO25”; not reported in this manuscript). Gray crosses (“ \times ”) indicate sessions for which MRI data was acquired but was excluded from all further analyses because it exceeded the head motion cutoff (>25% of fMRI volumes with >0.5 mm framewise displacement).

The structured intervention for the reading group focused on phonics (i.e., reading and writing the 46 primary Devanagari characters and their correspondence to sublexical consonant–vowel combinations), word decoding (i.e., reading and writing monosyllabic and more complex words), and sentence reading. The intervention was carried out by a local teacher and involved 2–4 hours of schooling on 5 days per week. Attendance of the classes was checked but not strictly enforced (mean = 3.74 days attended per week).

Experimental design

At each MRI scanning session, children performed an in-scanner language task with a block design consisting of short blocks of stimuli from six different conditions (visual words, auditory words, visual pseudowords, auditory pseudowords, visual low-level controls, and auditory low-level controls). All words were nouns of masculine grammatical gender, consisting of one or two syllables and three to six phonemes, and belonging to the same semantic category (animals). Pseudowords were generated from these words by replacing the initial consonant and vowel of the word. The substituted consonants were within the same articulatory place as the original consonants and the substituted graphemes matched the shape of the original graphemes as closely as possible. The low-level visual controls were false fonts that were created by rearranging the line segments of each grapheme of a word while preserving the position of the graphemes. The low-level auditory controls were created by spectrally rotating and noise-vocoding the spoken words. This was achieved by low-pass filtering the speech signal, inverting its spectrum around a center frequency of 2 kHz, dividing the speech signal into two logarithmically spaced frequency bands, extracting the amplitude envelope in each frequency band, using this envelope to modulate noise in the same frequency band, and recombining the frequency bands. All visual stimuli were presented in the middle of the screen in black font on a white background. The screen was placed behind the scanner bore and projected to the participant via a mirror mounted inside of the scanner. All auditory stimuli were recorded by a male native Hindi speaker.

In each of 108 blocks (18 per condition), 6 stimuli from the same condition were randomly presented with a duration of 1 s each. The order of blocks was random and not optimized for design efficiency. Between subsequent blocks, there was a random pause of 2.55, 3.82, or 5.09 s (equaling 1, 1.5, or 2 times the repetition time [TR]), during which a black fixation cross was presented in the middle of the screen. The total duration of the experiment was 17:40 min. To keep children attentive, they were asked to perform a simple target detection task. For this purpose, a target stimulus was inserted at a random location in 36 out of the 108 blocks. For visual blocks, this was the photograph of the face of a children's movie character, and for auditory blocks a short snippet of child-friendly human laughter. Children were asked to press a button on a MR-compatible button box whenever they saw or heard the target stimulus. They received auditory feedback in the form of a positive sound (after pressing the button when a target stimulus had appeared) or a negative sound (after not pressing the button when a target stimulus had appeared or after pressing the button when no target stimulus had appeared). The experiment was programmed and presented using PsychoPy (Version 2021) in Python.

Behavioral testing

At each session, either before or after MRI scanning, children completed a set of behavioral tests together with a

local research assistant. These tests were selected to assess children's reading skills, mathematics skills, working memory, and general cognitive ability.

For reading skills, we used the Middle Screening Tool (MST) version of the Dyslexia Assessment for Languages of India (DALI) test (Rao et al., 2015; Rao et al., 2021), with the following subtests:

- Rapid picture naming (time to name 50 object pictures)
- Word reading (number of correct Hindi words read and time taken [50 words])
- Rhyming (number of correctly identified rhyme pairs [12 items with three words each])
- Phoneme replacement (number of correct Hindi words generated by replacing one phoneme [10 items])
- Semantic fluency (number of Hindi words generated from two semantic categories [vegetables and animals] in one minute)
- Verbal fluency (number of Hindi words generated from two initial consonants in one minute)
- Pseudoword reading (number of correctly pronounced pseudowords and time taken [30 pseudowords])
- Reading comprehension (time taken to read a paragraph of text in Hindi + correct responses to comprehension questions [6 items])
- Dictation (number of Hindi words spelled correctly [20 words])

For mathematics skills, we used the Wide Range Achievement Test, Fifth Edition – India (WRAT5 – INDIA; Wilkinson & Robertson, 2017), with the following subtests:

- Oral math (number of questions answered correctly [15 items])
- Math computation (number of math problems solved correctly [40 items])

For working memory, we used the following tests:

- Corsi block-tapping test forward and backward (longest correctly repeated sequence; Corsi, 1972; Kessels et al., 2000)
- Digit span test forward and backward (longest correctly repeated sequence; Miller, 1956)

For general cognitive ability, we used Raven's Colored Progressive Matrices (CPM/CVS India; 36 items; Raven & Raven, 2003).

We analyzed the data from each behavioral (sub-)test with a linear mixed-effects model using the MixedModels package (Version 4.25.3; Bates et al., 2024) in Julia (Version 1.10.4; Bezanson et al., 2017). Each model predicted the raw (sub-)test scores using a fixed intercept (reflecting the average test score at the beginning of the study), a fixed effect of time (number of months since the beginning of the study), a by-participant random intercept and random slope for the effect of time, and their correlation. We

interpreted p values smaller than .05 for the fixed effect of time as a statistically significant increase (or decrease) in behavioral test performance.

MRI scanning parameters

All scanning was conducted on a GE SIGNA Architect 3T MRI machine with a 48 channel head coil. After a short head localizer scan, there was one functional scan, during which children performed the language experiment described above, and one structural scan, during which children watched a TV episode or video of their choice.

The functional scan was implemented using a gradient echo (GR) echo planar imaging (EPI) sequence with the following parameters: TE = 35 ms, TR = 2.547 s, flip angle = 88°, number of volumes = 420, field of view = 19.2 cm, in-plane matrix size = 80 × 80 voxels, slice thickness = 2.4 mm, gap between slices = 0.2 mm, voxel size = 2.4 × 2.4 × 2.6 mm, slice orientation = axial, phase encoding direction = anterior/posterior, slice order = interleaved/ascending, multiband acceleration (GE HyperBand) factor = 4. Slices covered the whole brain including the cerebellum. We did not collect any field maps and did not correct for potential inhomogeneity or spatial distortion.

The structural scan was implemented using a T1-weighted MPRAGE sequence with the following parameters: TE = 3.188 ms, TR = 2.30568 s, TI = 900 ms, flip angle = 8°, field of view = 22.4 cm, in-plane matrix size = 256 × 256 voxels, slice thickness = 0.9 mm, no gap between slices, voxel size = 0.875 × 0.875 × 0.9 mm, slice orientation = axial, phase encoding direction = right/left.

Preprocessing

Results included in this manuscript come from preprocessing performed using *fMRIprep* 24.0.1 (Esteban et al., 2018, RRID:SCR_016216; Esteban et al., 2019), which is based on *Nipype* 1.8.6 (Gorgolewski et al., 2011; Gorgolewski et al., 2018, RRID:SCR_002502).

Anatomical data preprocessing. A total of 2–6 T1-weighted (T1w) images were found within the input BIDS dataset. Each T1w image was corrected for intensity non-uniformity (INU) with *N4BiasFieldCorrection* (Tustison et al., 2010), distributed with ANTs 2.5.1 (Avants et al., 2008, RRID:SCR_004757). The T1w-reference was then skull-stripped with a *Nipype* implementation of the *antsBrainExtraction.sh* workflow (from ANTs), using OASIS30ANTS as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using *fast* (FSL, RRID:SCR_002823, Zhang et al., 2001). An anatomical T1w-reference map was computed after registration of 2–6 T1w images (after INU-correction) using *mri_robust_template* (FreeSurfer 7.3.2, Reuter et al., 2010). Brain surfaces were reconstructed using *recon-all* (FreeSurfer 7.3.2, RRID:SCR_001847, Dale et al., 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and

FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al., 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with *antsRegistration* (ANTs 2.5.1), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization and accessed with *TemplateFlow* (24.2.0, Ciric et al., 2022): *ICBM 152 Nonlinear Asymmetrical template version 2009c* (Fonov et al., 2009, RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym).

Functional data preprocessing. For each of the 2–6 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume was generated, using a custom methodology of *fMRIPrep*, for use in head motion correction. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using *mcflirt* (FSL, Jenkinson et al., 2002). The BOLD reference was then co-registered to the T1w reference using *bbregister* (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al., 2014) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al., 2002). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The

remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (*crown*) of voxels around the edge of the brain, as proposed by (Patriat et al., 2017). All resamplings can be performed with a *single interpolation step* by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `nitransforms`, configured with cubic B-spline interpolation. Non-gridded (surface) resamplings were performed using `mri_volsurf` (FreeSurfer).

Many internal operations of `fMRIPrep` use *Nilearn* 0.10.4 (Abraham et al., 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in `fMRIPrep`'s documentation (<https://fmriprep.readthedocs.io/en/latest/workflows.html>).

Copyright Waiver. The above boilerplate text was automatically generated by `fmriprep` with the express intention that users should copy and paste this text into their manuscripts *unchanged*. It is released under the CC0 license.

Additional preprocessing. After running `fMRIPrep`, the preprocessed BOLD fMRI time series data was spatially smoothed with a Gaussian kernel (FWHM = 5.0 mm) and masked using the whole-brain mask computed per-participant by `fMRIPrep`.

Session-level analysis

Separately for each participant and session, we modeled the preprocessed BOLD fMRI time series data using a mass-univariate general linear model (GLM) implemented in *Nilearn* (Version 0.10.4; Abraham et al., 2014) for Python (Version 3.9.19; Van Rossum & Drake, 2009). At each voxel, the BOLD time series (420 time points) was predicted using a design matrix with the following columns:

- A constant term (1 column).
- One task regressor for each of the six experimental conditions, created by convolving the block design (condition on/off) with the canonical “SPM” hemodynamic response function (HRF) implemented in *Nilearn* (6 columns). These were our regressors of interest.
- The first and second derivatives of each task regressor (12 columns), to capture age- and participant-specific deviations from the canonical HRF.
- The head motions parameters (translations and rotations in three directions) estimated by `fMRIPrep` during head motion correction (6 columns).

- Cosine regressors to high-pass filter the data at 128 s (≈ 0.008 Hz), removing slow-frequency scanner drifts (15 columns).
- The top six anatomical CompCor components (Behzadi et al., 2007) estimated by `fMRIPrep` (6 columns).
- Spike regressors for each non-steady state outlier volume at the beginning of the scan as flagged by `fMRIPrep` (0–3 columns, mean = 2.0 columns).
- Spike regressors for each high-motion outlier volume, defined as framewise displacement > 0.5 mm and flagged by `fMRIPrep` (0–103 columns, mean = 22.8 columns).

For details on how these regressors were computed, see the “Outputs/Confounds” section in `fMRIPrep`'s documentation (<https://fmriprep.org/en/stable/outputs.html#confounds>).

From the fitted model, we computed an effect size map (“beta map”) for each of the following contrasts, always reflecting the change in BOLD activation (in arbitrary units) between two experimental conditions:

- Auditory low-level vs. baseline
- Auditory pseudowords vs. baseline
- Auditory pseudowords vs. low-level
- Auditory words vs. baseline
- Auditory words vs. low-level
- Auditory words vs. pseudowords
- Visual low-level vs. baseline
- Visual pseudowords vs. baseline
- Visual pseudowords vs. low-level
- Visual words vs. baseline
- Visual words vs. low-level
- Visual words vs. pseudowords

We refer to all contrasts involving the baseline as “baseline contrasts,” which capture the difference in BOLD activity between each experimental condition and the fixation baseline period between experimental blocks. Therefore, these contrasts capture not only BOLD activity related to linguistic (phonological and semantic) processing but also BOLD activity related to low-level sensory processing (e.g., visual and auditory processing of the stimuli). We refer to all other contrasts as “experimental contrasts,” which capture the difference in BOLD activity between two different experimental conditions. The experimental contrasts comparing pseudowords to low-level controls capture BOLD activity related to phonological processing, since pseudowords but not low-level controls contain phonological information, while both do not contain any semantic information. The experimental contrasts comparing words to pseudowords capture BOLD activity related to semantic processing, since words but not pseudowords contain semantic information, while both contain phonological information. The experimental contrasts comparing words to low-level controls capture BOLD activity related to both phonological and semantic processing, since words but not low-level controls contain both phonological and semantic information.

Group-level analysis

BOLD activity amplitude. To estimate reading-related changes in BOLD activity amplitude, we fitted the beta maps from all participants and sessions using a linear mixed-effects model, separately for each contrast (see above). The dependent variable was the BOLD activation amplitude for a given participant and session, and the predictors were (1) a fixed intercept, implemented as a column of “1”s and reflecting the BOLD activity at the beginning of the study, (2) a fixed effect for linear time, implemented as the number of months elapsed since the beginning of the study and reflecting the linear change in BOLD activity due to the reading instruction, (3) a fixed effect for quadratic time, implemented as the square of linear time and reflecting the nonlinear (“u”-shaped or inverted “u”-shaped) change in BOLD activity due to the reading instruction, (4) a random intercept, reflecting individual participant’s deviation from the global BOLD activity at the beginning of the study, and random slopes for (5) linear and (6) quadratic time, reflecting individual participant’s deviations in the linear and nonlinear changes in BOLD activity due to the reading instruction. As is typical in frequentist linear mixed models, one parameter was estimated for each fixed effect (intercept, linear time, and quadratic time) and for each random effect (the standard deviation of the random intercepts and slopes for linear and quadratic time), as well as three correlation parameters between the three pairs of random effects. The mixed models (one for each contrast) were fitted separately at each voxel inside the brain mask in a mass-univariate fashion. Model fitting was performed in Julia (Version 1.10.4; Bezanson et al., 2017) using the MixedModels package (Version 4.25.3; Bates et al., 2024).

To correct for multiple comparisons across the 132,215 voxels inside the brain mask, we used the parametric cluster correction algorithm suggested by Cox et al. (2017a) and Cox et al. (2017b). Specifically, we first estimated the spatial smoothness of noise in our dataset by extracting and storing the residual time series from the session-level models (see above), separately for each participant and session. We then used the `3dFWHMx` program (with the `-acf` option) in AFNI (Version 24.2.01; Cox, 1996) to estimate a mixed Gaussian/mono-exponential spatial autocorrelation function with three parameters (Cox et al., 2017b; Cox et al., 2017a; see also https://afni.nimh.nih.gov/pub/dist/edu/data/CD.expanded/afni_handouts/afni07_ETAC.pdf). We averaged each of these three parameters across participants and sessions to obtain a single spatial autocorrelation function for the entire dataset. This function was then fed into the `3dClustSim` program in AFNI to generate novel noise-only maps and estimate a null distribution of cluster sizes. Using a cluster-forming voxel-level threshold of $p < .001$ and 10,000 iterations, this resulted in a final cluster-level extent threshold of 19 voxels to control the whole-brain family-wise error (FWE) rate at $p < .05$. We therefore deemed spatial clusters of BOLD activation statistically significant if they were larger than 19 voxels. To form clusters, neighboring voxels had to pass the cluster-forming voxel level threshold of $p < .001$ and touch each other with their faces (not just with their edges or nodes; the default “NN1” method in AFNI).

BOLD activity patterns. We also estimated reading related changes in the between-condition similarity and within-condition stability of BOLD activity patterns using multivariate analysis methods.

First, we investigated if the reading intervention made written word activity patterns more similar to those for spoken word activity patterns. For this, we used the beta maps from these two conditions and extracted the betas for different regions of interest (ROI; defined below). We then computed the linear correlation between these vectors and entered these correlations (one value per subject and session) into a linear mixed-effects model, separately for each ROI. We specified the linear mixed-effects model in the same way as described above, with fixed and random effects for the intercept, the linear effect of time, and the quadratic effect of time. We repeated the same analysis for the correlation of activity patterns between written and spoken pseudowords and between written and spoken low-level controls (both baseline contrasts and experimental contrasts). However, it is important to note at this point that with our block design, we are not comparing the activity patterns in response to individual items (e.g., the written word “monkey” and the spoken word “monkey”), but rather the general patterns of activity when processing written and spoken words. Though arguably more meaningful, the former comparison was not possible because our stimuli within each block were presented faster (1 s) than our TR (2.647 s), and we were therefore unable to obtain a beta map for each individual stimulus.

Second, we investigated if the reading intervention made written word activity more “stable,” i.e., more consistent across repeated presentations of written words within the same session. For this, we re-estimated the session-level beta maps as described above but with separate task regressors for each block (108 columns) instead of each condition (6 columns). We then took the beta map for each written word block (18 maps) and extracted the betas for different ROIs (defined below). We then computed the linear correlation between each pair of blocks (153 pairs) and averaged these to obtain a single stability (correlation) value for each participant and session. We entered these values into a linear-mixed effects model, separately for each ROI and specified in the same way as described above. We repeated the same analysis for the stability of activity patterns for written pseudowords and written low-level controls. We also performed this analysis for spoken words, spoken pseudowords, and spoken low-level controls, even though they do not directly pertain to our hypotheses.

For all multivariate analyses, we used the same set of anatomically and functionally defined regions of interest. The anatomically defined regions of interest were the posterior superior temporal sulcus (pSTS), defined as the union of regions STSda, STSdp, STSvp, and STSva from the Glasser et al. (2016) atlas, and the ventral occipito-temporal cortex (vOT), defined as the union of regions V8, FFC, and VVC from the Glasser et al. (2016) atlas. The functionally defined regions were different for each condition: For the similarity between written and spoken words, we used the visual clusters from the “Written words vs. baseline” contrast and the auditory

clusters from the “Spoken words vs. baseline” clusters. Likewise, for the similarity between written and spoken pseudowords, we used the visual clusters from the “Written pseudowords vs. baseline” contrast and the auditory clusters from the “Spoken pseudowords vs. baseline” clusters. For the within-condition pattern stability analysis, we used the visual or auditory clusters from that specific condition (e.g., for the stability analysis for written words, we used the visual clusters from the “Written words vs. baseline” contrast).

Data and code availability

The data from this study are available upon request from the last author of the paper. The code for the experiment, preprocessing, and statistical analysis is available at <https://github.com/SkeideLab/SLANG>.

Results

Behavioral testing

At each session throughout the study, children completed a set of behavioral tests to assess their reading skills, mathematics skills, working memory, and general cognitive ability (see Figure 2).

The following (sub-)tests showed a significant change in test scores over the course of the study:

- Word reading accuracy ($b = 0.42$ more correct words per month, $p = .025$)
- Phoneme replacement ($b = 0.18$ more items correct per month, $p = 0.011$)
- Semantic fluency ($b = 0.26$ more generated words per month, $p = 0.015$)
- Pseudoword reading ($b = 0.20$ more correct pseudowords per month, $p = 0.042$)
- Reading comprehension ($b = 0.06$ more questions answered correctly per month, $p = 0.042$)
- Dictation ($b = 0.27$ more correctly spelled words per month, $p = 0.032$)
- Oral math ($b = 0.07$ more items solved correctly per month, $p = 0.017$)
- Math computation ($b = 0.30$ more items solved correctly per month, $p = 0.018$)
- Digit span forward ($b = 0.08$ more digits per month, $p = 0.017$)

All other (sub-)tests (picture naming, word reading time, rhyming, verbal fluency, pseudoword reading time, passage reading time, Corsi block forward and backward, digit span backward, Raven’s Colored Progressive Matrices) showed no statistically significant change over the course of the study (all $ps > .072$). For all (sub-)tests, there was large interindividual variation between children, both in their initial scores as well as in their longitudinal change over time (see Figure 2).

BOLD activity amplitude

Auditory baseline contrasts. At the beginning of the study (intercept), spoken low-level controls, spoken pseudowords, and spoken words elicited BOLD activity in the left and right auditory cortex (superior temporal gyrus) and in the left inferior frontal gyrus (see orange clusters in Figures 3A, 4A, and 5A), as well as a few BOLD deactivations in parietal and occipital areas (see blue clusters in Figures 3A, 4A, and 5A). For spoken low-level controls, there was positive linear change in BOLD activity over the course of the study in one small cluster near the right motor cortex (see orange cluster in Figure 3C) and positive quadratic (i.e., “u”-shaped) change in BOLD activity three small clusters in the right middle/inferior temporal lobe and in the right parietal lobe (see orange clusters in Figure 3E). For spoken pseudowords, there was both negative linear and positive quadratic (i.e., “u”-shaped) change in one small cluster in the right parietal lobe (see blue cluster in Figure 4C and orange cluster in Figure 4E). For spoken words, there was positive linear change in BOLD activity over the course of the study in one small cluster in the left inferior parietal lobe (see orange cluster in Figure 5C), as well as positive quadratic change (i.e., “u”-shaped) change in one small cluster in the right superior parietal lobe (see orange cluster in Figure 5E).

Visual baseline contrasts. At the beginning of the study (intercept), written low-level controls, written pseudowords, and written words elicited BOLD activity in the left and right occipital and inferior temporal cortices (see orange clusters in Figures 3B, 4B, and 5B), as well as BOLD deactivations in the left and right medial occipital lobes (see blue clusters in Figures 3B, 4B, and 5B). For written low-level controls, there was no linear change in BOLD activity over the course of the study (see Figure 3D) but positive quadratic (i.e., “u”-shaped) change in BOLD activity two small clusters in the right occipital lobe (see orange clusters in Figure 3F). For written pseudowords, there was negative linear change in one small cluster in the left inferior frontal lobe (see left frontal blue cluster in Figure 4D) and both negative linear change and positive quadratic (i.e., “u”-shaped) change in one small cluster in the right occipital lobe (see posterior blue cluster in Figure 4D and orange cluster in Figure 4F). For written words, there was both negative linear and positive quadratic (i.e., “u”-shaped) change in one small cluster in the right occipital lobe (see small blue cluster in Figure 5D and posterior orange cluster in Figure 5F). Additionally, there was positive quadratic (i.e., “u”-shaped) change in one small cluster in the right parietal lobe (see anterior orange cluster in Figure 5F).

Auditory experimental contrasts. At the beginning of the study (intercept), the difference between spoken pseudowords and spoken low-level controls elicited widespread BOLD activity mainly in the left and right auditory cortex (superior temporal gyrus; see large orange clusters in Figure 6A). There was negative linear change in BOLD activity over the course of the study in one small cluster in the right anterior inferior parietal lobe (see small blue cluster in Figure 6C). There was positive quadratic change in BOLD activity over the course of the

10

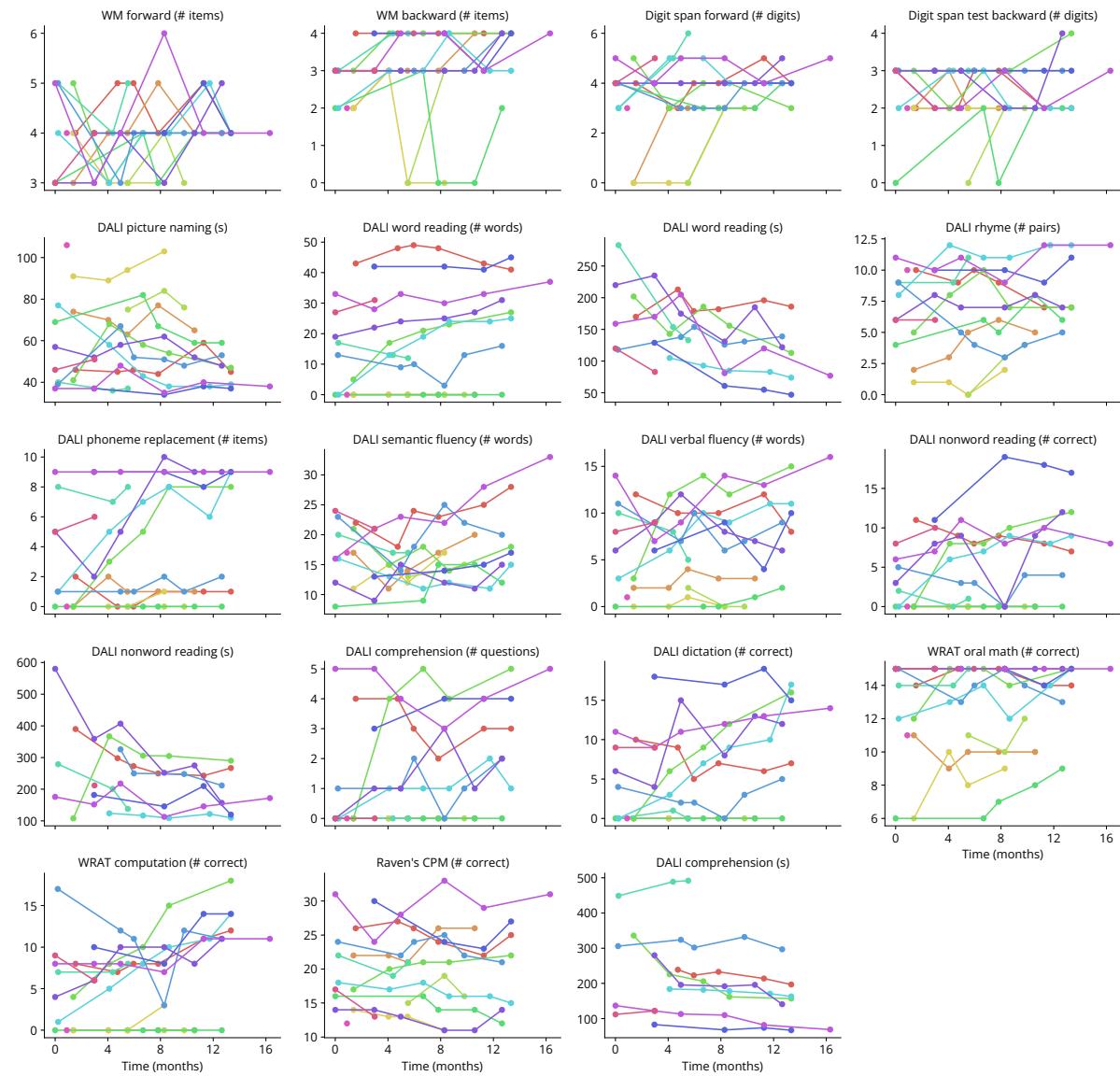
ALEXANDER ENGE^{1,2} & MICHAEL A. SKEIDE¹

Figure 2. Results from behavioral tests of reading skills, mathematics skills, working memory, and general cognitive ability. Each subplot shows the results (raw scores) from one subtest. Colored dots and lines show the results from individual children over the course of the study. Note that for subtests where the score is the number (#) of correct responses, higher values are considered better, whereas for subtests where the score is the number of seconds (s) taken to complete the test, lower values are considered better. DALI = Dyslexia Assessment for Languages of India, WRAT = Wide Range Achievement Test, WM = working memory (Corsi block test), CPM = Raven's Colored Progressive Matrices.

study in one small area in the left superior parietal lobe (see small orange cluster in Figure 6E).

At the beginning of the study (intercept), the difference between spoken words and spoken low-level controls elicited widespread BOLD activity mainly in the left and right auditory cortex (superior temporal gyrus; see large orange clusters in Figures 7A). There was no evidence for linear or quadratic change over the course of the study in any brain areas (see Figure 7C and E).

The difference between spoken words and spoken pseudowords did not elicit any reliable BOLD activity at the beginning of the study and showed no evidence for change over the course of the study (see Figure 8A, C, and E).

Visual experimental contrasts. At the beginning of the study (intercept), the difference between writ-

ten pseudowords and written low-level controls elicited BOLD activity in one cluster in the right dorso-lateral prefrontal cortex (see orange cluster in Figure 6B) and BOLD deactivation in one cluster in the right secondary visual cortex (see blue cluster in Figure 6B). There was negative linear change over the course of the study in the right dorso-lateral prefrontal cortex (see blue cluster in Figure 6D).

The difference between written words and written low-level controls did not elicit any reliable BOLD activity at the beginning of the study (see Figure 7B) but there was negative linear change over the course of the study in one cluster in the left supplementary motor area (see small blue cluster in Figure 7D) and negative quadratic (inverted "u"-shaped) change over the course of the study in one cluster in the left posterior medial wall (near the

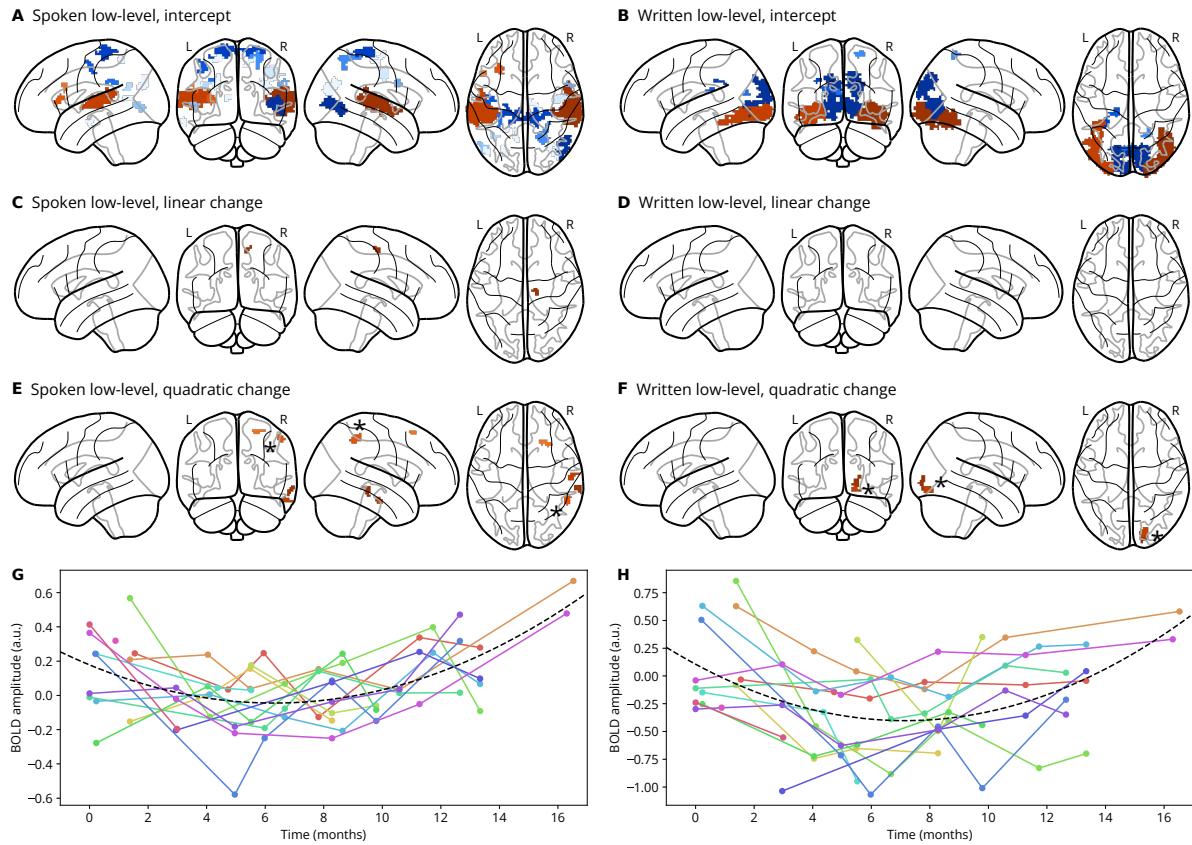


Figure 3. BOLD activity in response to low-level sensory controls. Panels A and B show statistically significant clusters (cluster-forming voxel threshold $p < .001$, uncorrected; cluster size threshold $p < .05$, FWE-corrected) for the contrast of low-level sensory control blocks vs. fixation baseline at the beginning of the study (intercept; time = 0 months). Panels C and D show statistically significant clusters for the linear change in BOLD activity for the same contrast over the course of the study and panels E and F show statistically significant clusters for the quadratic change in BOLD activity for the same contrast over the course of the study. Panels A, C, and E show the auditory modality (noise-vocoded speech vs. fixation baseline) and panels B, D, and F show the visual modality (false fonts vs. fixation baseline). In all panels A–F, clusters with positive BOLD amplitude (low-level > baseline) are shown in orange and clusters with negative BOLD amplitude (low-level < baseline) are shown in blue. Clusters with higher voxel-level peak statistics are shown in brighter colors. Panels G and H show individual participants' change in BOLD amplitude over time (colored dots and lines) as well as the best fitting linear mixed model (dashed black line) for the largest significant clusters in panels E and F, respectively, as indicated by the black asterisk (*) next to the cluster.

ventral posterior cingulate cortex; see the blue cluster in Figure 7F).

The difference between written words and written pseudowords elicited BOLD deactivation in one cluster in the right dorso-lateral prefrontal cortex (see blue cluster in Figure 8B). There was no evidence for linear or quadratic change over the course of the study (see Figure 8D and F).

BOLD activity patterns

Audio-visual pattern similarity. For all baseline contrasts (low-level vs. baseline, pseudowords vs. baseline, and words vs. baseline) and ROIs, the auditory and visual BOLD response patterns were correlated significantly at the beginning of the study (intercept; all $p < .001$; see Figures 9–11). This is expected given that the exact same fixation baseline periods were used as the comparison condition for both the auditory and visual baseline contrasts. There was no evidence for linear change (all $p > .072$) in pattern similarity over the course of the study for any contrast pair or ROI.

For all experimental contrasts (pseudowords vs. low-level, words vs. low-level, words vs. pseudowords) and ROIs,

the auditory and visual BOLD response patterns were not significantly correlated at the beginning of the study (intercept; all $p > .090$; see Figures 12–14). There was a linear increase in audio-visual pattern similarity over the course of the study for the contrast of words versus low-level controls in the left ventral occipito-temporal (vOT) ROI ($b = 0.013$, $p = 0.034$, see Figure 13C). There was no evidence for linear change in audio-visual pattern similarity over the course of the study for any of the other experimental contrasts and ROIs (all $p > .121$; see Figures 12–14).

Within-condition pattern stability. For almost all conditions and ROIs, blocks from the same condition were positively correlated at the beginning of the study (intercept; all $p < .029$; see Figures 15–20), except for written low-level controls in the left spoken low-level controls ROI ($p = 0.570$, see Figure 16E), for written pseudowords in the left spoken pseudoword ROI ($p = 0.346$, see Figure 18E) and in the right spoken pseudoword ROI ($p = 0.084$, see Figure 18F), and for written words in the left spoken word ROI ($p = 0.201$, see Figure 20E). There was a decrease in BOLD pattern stability for spoken pseudowords in the left pSTS ROI ($b = -0.003$, $p = 0.029$, see Figure 17A) and in the right pSTS ROI ($b = -0.003$,

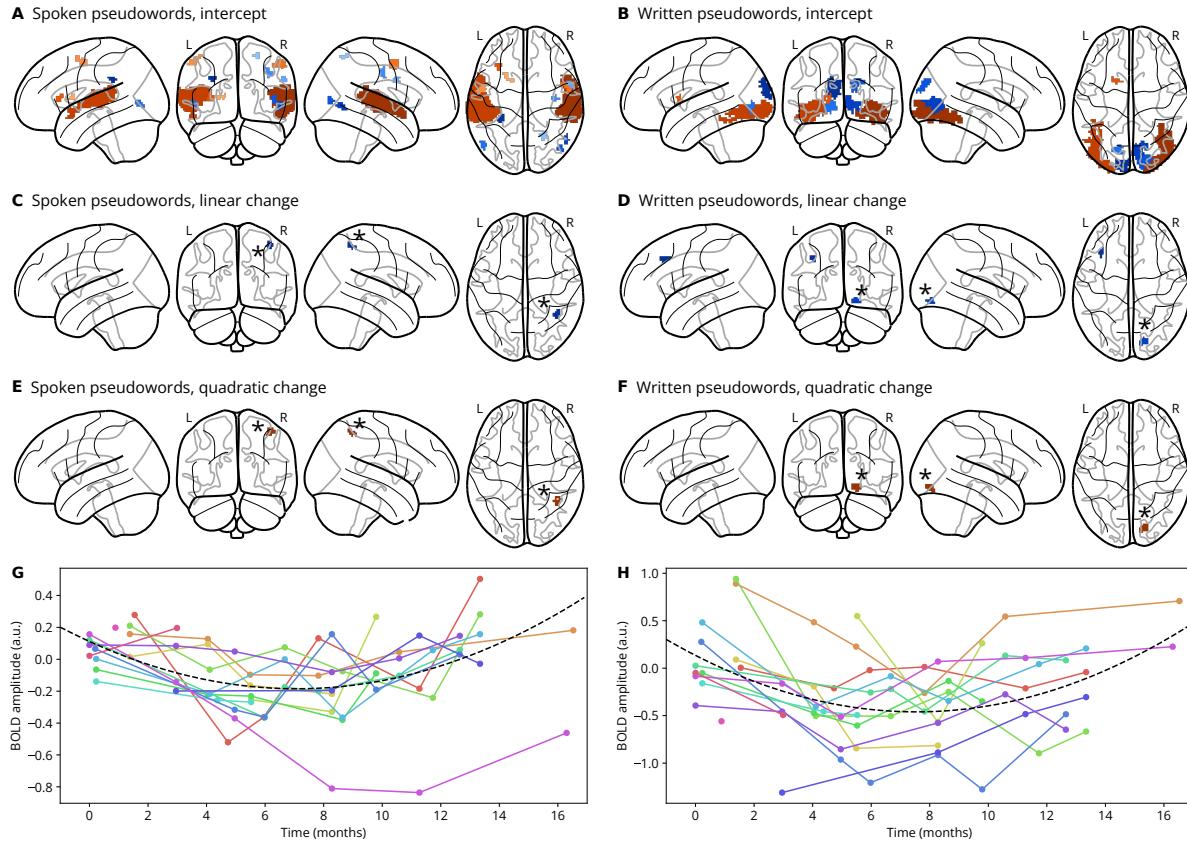


Figure 4. BOLD activity in response to pseudowords. Panels A and B show statistically significant clusters (cluster-forming voxel threshold $p < .001$, uncorrected; cluster size threshold $p < .05$, FWE-corrected) for the contrast of pseudoword blocks vs. fixation baseline at the beginning of the study (intercept; time = 0 months). Panels C and D show statistically significant clusters for the linear change in BOLD activity for the same contrast over the course of the study and panels E and F show statistically significant clusters for the quadratic change in BOLD activity for the same contrast over the course of the study. Panels A, C, and E show the auditory modality (spoken pseudowords vs. fixation baseline) and panels B, D, and F show the visual modality (written pseudowords vs. fixation baseline). In all panels A–F, clusters with positive BOLD amplitude (pseudowords > baseline) are shown in orange and clusters with negative BOLD amplitude (pseudowords < baseline) are shown in blue. Clusters with higher voxel-level peak statistics are shown in brighter colors. Panels G and H show individual participants' change in BOLD amplitude over time (colored dots and lines) as well as the best fitting linear mixed model (dashed black line) for the largest significant clusters in panels E and F, respectively, as indicated by the black asterisk (*) next to the cluster.

$p = 0.039$, see Figure 17B), as well as for spoken words in the left pSTS ROI ($b = -0.002$, $p = 0.040$, see Figure 19A), in the right pSTS ROI ($b = -0.004$, $p = 0.013$, see Figure 19B), and in the right spoken words ROI ($b = -0.006$, $p = 0.046$, see Figure 19F). Note that we did not have any *a priori* hypothesis for longitudinal changes in pattern stability for these auditory conditions but instead expected changes in pattern stability in the written pseudoword and word conditions, which we did not observe here.

Discussion

We used dense-sampling fMRI to longitudinally measure changes in children's brain activity as they were learning to read. Fifteen children received reading instruction in their native language (Hindi) and writing system (Devanagari) for approximately 16 months, during which they also participated in up to 6 fMRI scanning and behavioral testing sessions. In this sample, we found (a) some evidence for an improvement in reading performance based on standardized reading skill tests, (b) limited evidence for longitudinal increases (linear and non-linear) in BOLD activity in response to spoken and

written words, and (c) limited evidence for longitudinal increases in word-related audiovisual BOLD pattern similarity. We will discuss each of these findings in turn as well as the limitations and research implications of the present study.

Changes in behavioral test scores

As would be expected from approximately 1.5 years of systematic reading instruction, children's test scores on various aspects of reading performance improved over the course of the study. This included correctly reading real Hindi words and pseudowords as well as language skills closely related to reading, such as phoneme replacement and semantic fluency. We additionally observed improvements in some non-reading related tests (maths and working memory), which may be the consequence of general cognitive development and/or informal schooling from parents or siblings.

Non-linear change in BOLD activity amplitude

We had hypothesized that learning to read would lead to an increase in pseudoword- and word-related BOLD ac-

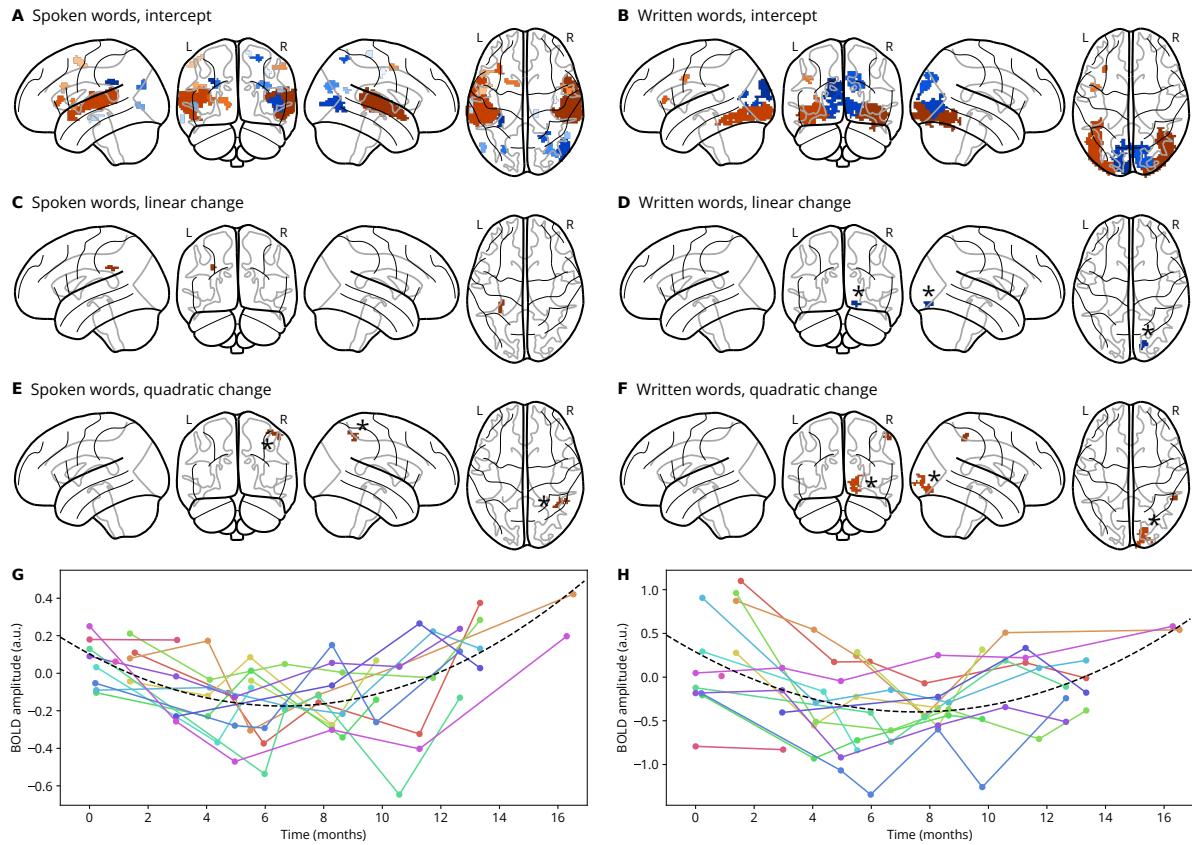


Figure 5. BOLD activity in response to words. Panels A and B show statistically significant clusters (cluster-forming voxel threshold $p < .001$, uncorrected; cluster size threshold $p < .05$, FWE-corrected) for the contrast of word blocks vs. fixation baseline at the beginning of the study (intercept; time = 0 months). Panels C and D show statistically significant clusters for the linear change in BOLD activity for the same contrast over the course of the study and panels E and F show statistically significant clusters for the quadratic change in BOLD activity for the same contrast over the course of the study. Panels A, C, and E show the auditory modality (spoken words vs. fixation baseline) and panels B, D, and F show the visual modality (written words vs. fixation baseline). In all panels A–F, clusters with positive BOLD amplitude (words > baseline) are shown in orange and clusters with negative BOLD amplitude (words < baseline) are shown in blue. Clusters with higher voxel-level peak statistics are shown in brighter colors. Panels G and H show individual participants' change in BOLD amplitude over time (colored dots and lines) as well as the best fitting linear mixed model (dashed black line) for the largest significant clusters in panels E and F, respectively, as indicated by the black asterisk (*) next to the cluster.

tivity amplitude in areas that are associated with reading (e.g., the VWFA in the vOT cortex) and audiovisual integration (e.g., the pSTS). We had expected this change to be either linear, that is, increasing from the beginning to the end of the study, or negative quadratic (inverted “u”-shaped), that is, increasing in the beginning of the study and then decreasing towards the end of the study. We did not observe such a pattern for any experimental contrast or brain region, except for the contrast between written words and false fonts (see Figure 7). However, this effect was located at the medial wall of the brain (next to the posterior cingulate cortex), an area that has not typically been described as relevant for reading. For some of the other contrasts, a few areas showed evidence for linear and/or quadratic change over the course of the study, but this change typically did not follow the expected pattern (e.g., negative linear change or positive quadratic change) and did not occur in areas that are relevant for reading or audiovisual integration. Most importantly, most effects were observed for the baseline contrasts, contrasting one experimental condition against the fixation baseline (see Figures 3 to 5). These contrasts are difficult to interpret as they capture not only higher-level (e.g., orthographic, phonological, and semantic) processing, but also low-level sensory processing. The experimental con-

trasts (contrasting word-like stimuli with different levels of phonological and semantic information, e.g., written words vs. pseudowords) showed little to no reliable longitudinal change in BOLD activity (see Figures 6 to 8). There are several factors that could explain these null effects, including the low sample size in our study, an intervention that was too short or too weak, or a lack of sufficient high-quality data per participant and session. These limitations will be discussed in more detail further below.

Increase in audio-visual pattern similarity

We had hypothesized that learning to read would lead BOLD activity patterns to become more similar in response to written and spoken (pseudo-)words, especially in areas associated with reading (e.g., the VWFA in the vOT cortex) and audiovisual integration (e.g., the pSTS). Indeed, we did find an increase in audio-visual multi-voxel BOLD pattern similarity: Responses for written words (versus written low-level controls) became more similar over time to the responses to spoken words (versus spoken low-level controls) in the left ventral occipito-temporal cortex (see Figures 13C). However, beyond the

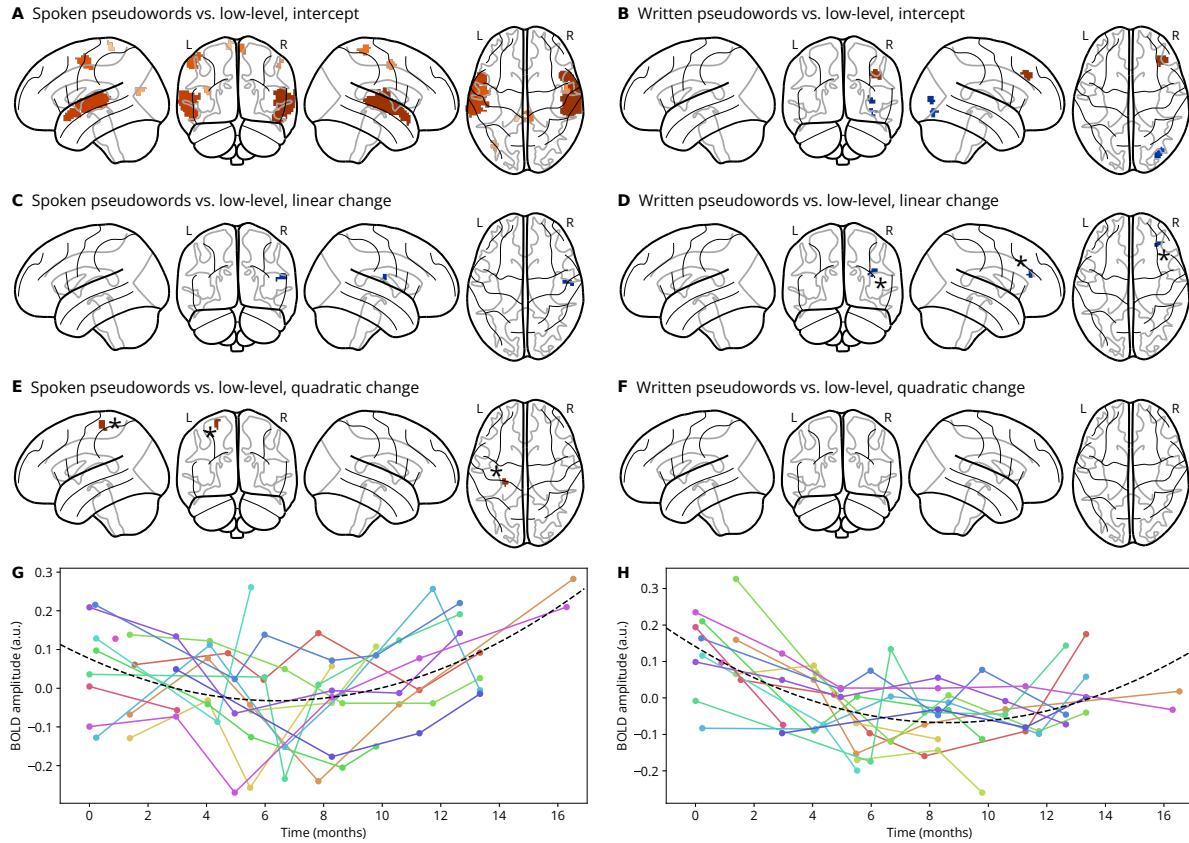


Figure 6. BOLD activity in response to pseudowords vs. low-level controls. Panels A and B show statistically significant clusters (cluster-forming voxel threshold $p < .001$, uncorrected; cluster size threshold $p < .05$, FWE-corrected) for the contrast of pseudoword blocks vs. low-level control blocks at the beginning of the study (intercept; time = 0 months). Panels C and D show statistically significant clusters for the linear change in BOLD activity for the same contrast over the course of the study and panels E and F show statistically significant clusters for the quadratic change in BOLD activity for the same contrast over the course of the study. Panels A, C, and E show the auditory modality (spoken pseudowords vs. noise-vocoded speech) and panels B, D, and F show the visual modality (written pseudowords vs. false fonts). In all panels A–F, clusters with positive BOLD amplitude (pseudowords > low-level) are shown in orange and clusters with negative BOLD amplitude (pseudowords < low-level) are shown in blue. Clusters with higher voxel-level peak statistics are shown in brighter colors. Panels G and H show individual participants' change in BOLD amplitude over time (colored dots and lines) as well as the best fitting linear mixed model (dashed black line) for the largest significant clusters in panels E and F, respectively, as indicated by the black asterisk (*) next to the cluster.

limitations mentioned above (and discussed in detail below), it is important to mention that our study design was not ideal for capturing similarities in BOLD activity patterns. This is because we had used a block design which did not allow us to estimate activity patterns for individual stimuli (e.g., the written word “monkey”) but only for experimental conditions (i.e., written words in general). Assuming that different stimuli elicit different BOLD activity patterns (e.g., depending on their individual phonology and semantics), we would have needed to use an event-related design to accurately capture these patterns and their potential increase in similarity between different sensory input modalities. However, it may also be that our initial hypothesis was wrong and that written and spoken words are processed in completely separate processing streams, converging only at some very abstract semantic level (likely situated in a region that was not captured by any of our ROIs). While audio-visual integration regions like the pSTS have been shown to respond strongly to input from both modalities (e.g., Blau et al., 2010; Calvert et al., 1997; van Atteveldt et al., 2004; Wilson et al., 2018), they may not be engaged in tasks where stimuli are presented in only one modality at a time (either spoken or written), and where they

do not need to be evaluated for audio-visual congruence. Finally, it may also be that automatic audio-visual processing of written words only develops relatively late after starting to learn to read, and that this developmental change therefore was not captured during the duration of our study.

Finally, we also tested for longitudinal increases in multi-voxel BOLD response pattern stability, that is, if response patterns to repeated presentations of the same category of stimuli (e.g., written words) became more similar to one another as children learnt to read. We did find some evidence for longitudinal changes in pattern stability, but only for spoken pseudowords and spoken words in the bilateral STS. Of note, pattern stability for these conditions and areas *decreased* rather than *increased* over the course of the study. Since we did not predict any change in pattern stability for these auditory conditions, we refrain from interpreting these findings.

Limitations

Almost every empirical research study has limitations and this one is definitely no exception. First and fore-

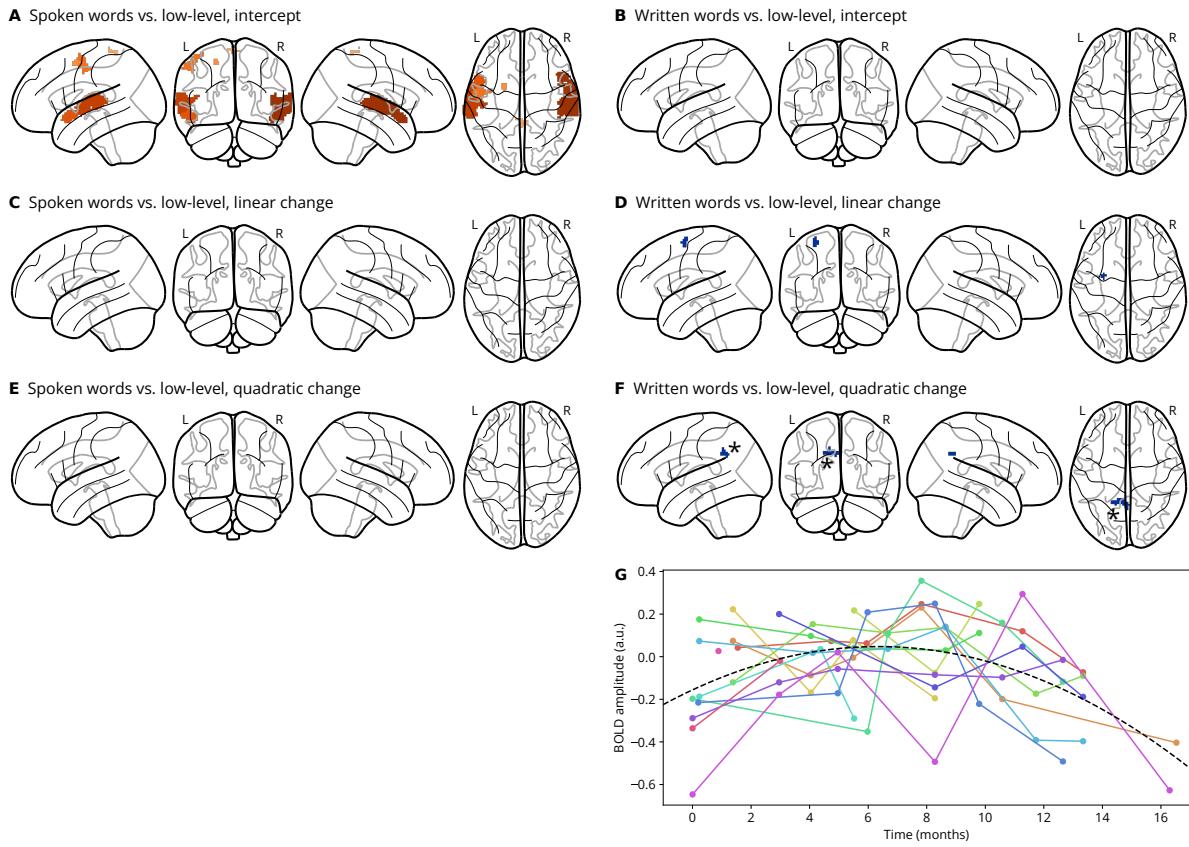


Figure 7. BOLD activity in response to words vs. low-level controls. Panels A and B show statistically significant clusters (cluster-forming voxel threshold $p < .001$, uncorrected; cluster size threshold $p < .05$, FWE-corrected) for the contrast of word blocks vs. low-level control blocks at the beginning of the study (intercept; time = 0 months). Panels C and D show statistically significant clusters for the linear change in BOLD activity for the same contrast over the course of the study and panels E and F show statistically significant clusters for the quadratic change in BOLD activity for the same contrast over the course of the study. Panels A, C, and E show the auditory modality (spoken words vs. noise-vocoded speech) and panels B, D, and F show the visual modality (written words vs. false fonts). In all panels A–F, clusters with positive BOLD amplitude (words > low-level) are shown in orange and clusters with negative BOLD amplitude (words < low-level) are shown in blue. Clusters with higher voxel-level peak statistics are shown in brighter colors. Panel G shows individual participants' change in BOLD amplitude over time (colored dots and lines) as well as the best fitting linear mixed model (dashed black line) for the largest significant cluster in panel F, as indicated by the black asterisk (*) next to the cluster.

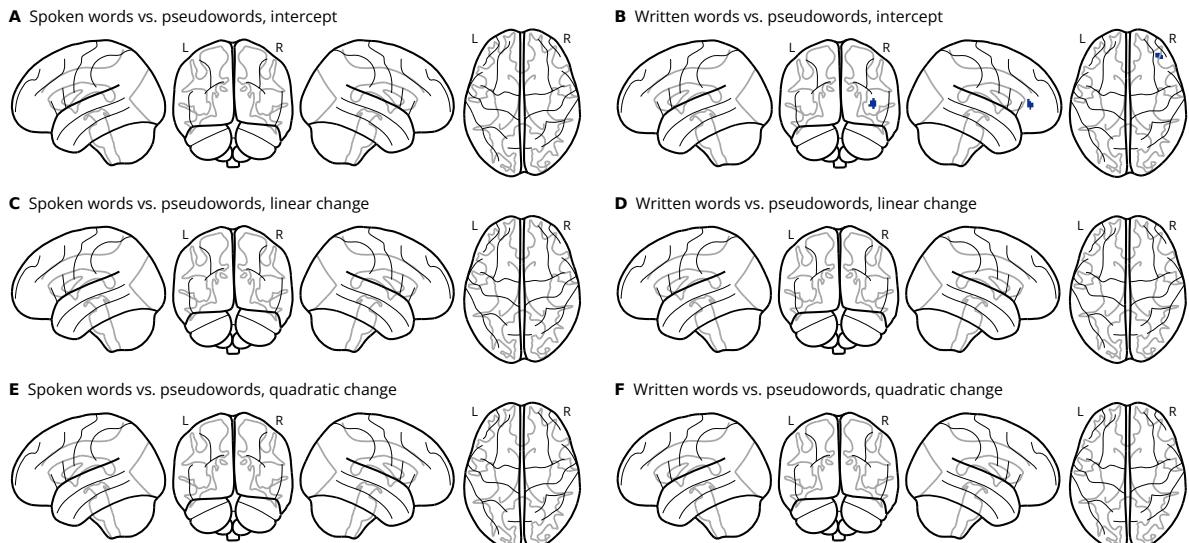


Figure 8. BOLD activity in response to words vs. pseudowords. Panels A and B show statistically significant clusters (cluster-forming voxel threshold $p < .001$, uncorrected; cluster size threshold $p < .05$, FWE-corrected) for the contrast of word blocks vs. pseudoword blocks at the beginning of the study (intercept; time = 0 months). Panels C and D show statistically significant clusters for the linear change in BOLD activity for the same contrast over the course of the study and panels E and F show statistically significant clusters for the quadratic change in BOLD activity for the same contrast over the course of the study. Panels A, C, and E show the auditory modality (spoken words vs. spoken pseudowords) and panels B, D, and F show the visual modality (written words vs. written pseudowords). In all panels A–F, clusters with positive BOLD amplitude (words > pseudowords) are shown in orange and clusters with negative BOLD amplitude (words < pseudowords) are shown in blue. Clusters with higher voxel-level peak statistics are shown in brighter colors.

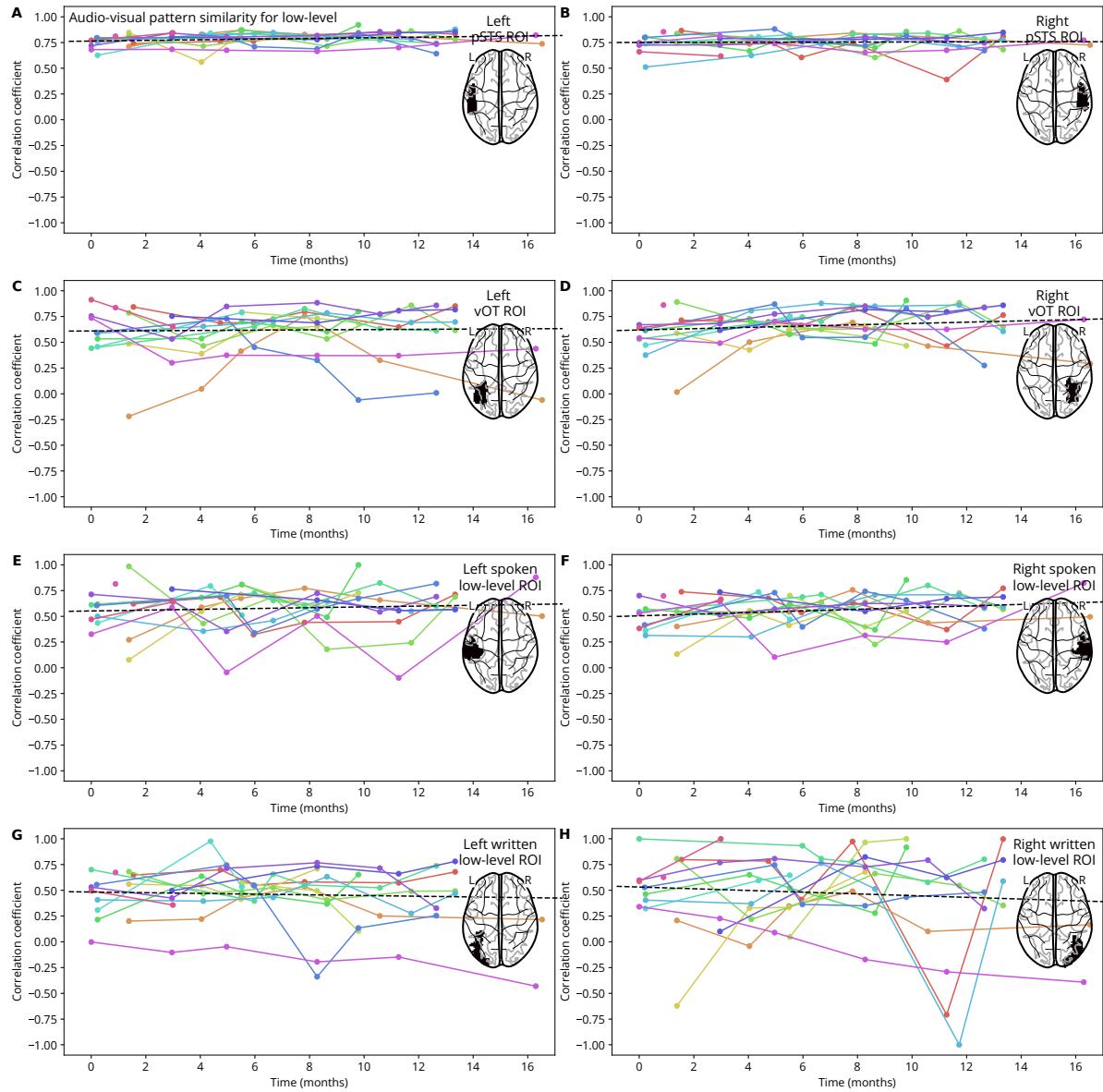


Figure 9. Audio-visual pattern similarity for low-level controls. Each panel A–H shows the development of audio-visual pattern similarity for the contrast of low-level sensory control blocks vs. fixation baseline over the course of the study in one region of interest (ROI). Audio-visual pattern similarity was computed by correlating the beta weights of all voxels inside the ROI for the auditory contrast (noise-vocoded speech vs. fixation baseline) with those for the visual condition (false fonts vs. fixation baseline). ROIs were defined either anatomically (posterior superior temporal sulcus [pSTS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 3). Colored dots and lines indicate the pattern similarities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

most, the sample size of this study was extremely small, both in terms of the number of participants ($N = 15$ in the reading intervention group) and in terms of the number of longitudinal sessions per participant ($T = 6$ for most participants). With this small sample size, statistical power for detecting effects is very low, which can lead to (a) undetected true effects (false negatives), (b) potentially unreplicable obtained effects (low positive predictive value), and (c) potentially inflated effect sizes for obtained effects (Button et al., 2013; Ioannidis, 2005; Szucs & Ioannidis, 2017). Unfortunately, BOLD signal changes for higher-level (i.e., non-sensory) experimental manipulations (e.g., words vs. pseudowords) are typically subtle in size compared to the neural, thermal, and scanner-induced noise

of fMRI data. When wanting to track changes longitudinally from session to session, the effects of interest can be presumed to be even smaller (since one is interested in the manipulation \times time interaction rather than in the main effect of the manipulation). These problems of small sample size, small effect size, and low statistical power are potentiated when scanning children because their short attention span and high head motion lead to both shorter scanning sessions (i.e., less trials/volumes) and lower data quality (i.e., higher framewise displacement and signal intensity changes). To reliably answer the research questions posed in the Introduction, one would presumably need a sample size at least 10 times larger than ours in terms of participants and ideally also significantly larger

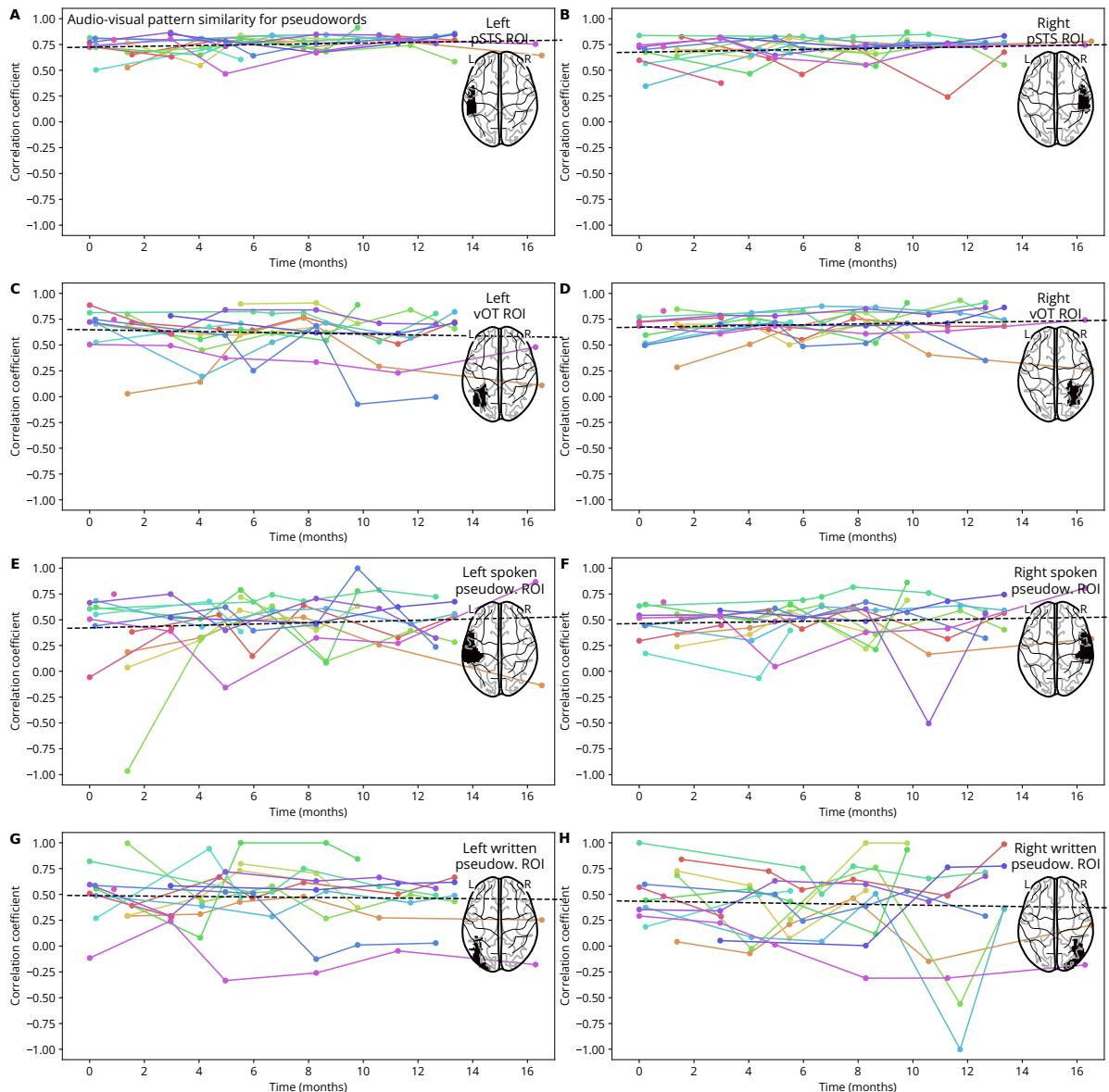


Figure 10. Audio-visual pattern similarity for pseudowords. Each panel A–H shows the development of audio-visual pattern similarity for the contrast of pseudoword blocks vs. fixation baseline over the course of the study in one region of interest (ROI). Audio-visual pattern similarity was computed by correlating the beta weights of all voxels inside the ROI for the auditory contrast (spoken pseudowords vs. fixation baseline) with those for the visual condition (written pseudowords vs. fixation baseline). ROIs were defined either anatomically (posterior superior temporal sulcus [STS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 4). Colored dots and lines indicate the pattern similarities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

in terms of the number of sessions per participant, both of which we failed to acquire due to monetary and time constraints.

Second, the reading intervention employed in the present study might not have been very effective, as we were unable to closely monitor the quantity and quality of the teaching sessions. While the behavioral test scores (see Figure 2) indicated some improvement on relevant subtests of reading ability for some children, we by and large cannot be sure how many children became significantly better readers over the course of the study. Therefore, it is not surprising that we found little to no reliable change in word-related BOLD activity amplitudes and patterns. On a similar note, the duration of our intervention (ap-

prox. 16 months) might not have been long enough for children to become sufficiently proficient readers, especially since the Devanagari script contains a relatively large number of visually complex letters. Our study design therefore was not ideal to capture the full process of learning to read and might have missed especially the late, *orthographic* stage of reading development (see Introduction).

Finally, there were minor issues in our experimental design that could have been improved, including (a) a relatively long fMRI scanning run, which might have induced tiredness and additional head motion for some participants, (b) relatively short baseline intervals between blocks, which might have led to overlap of BOLD activi-

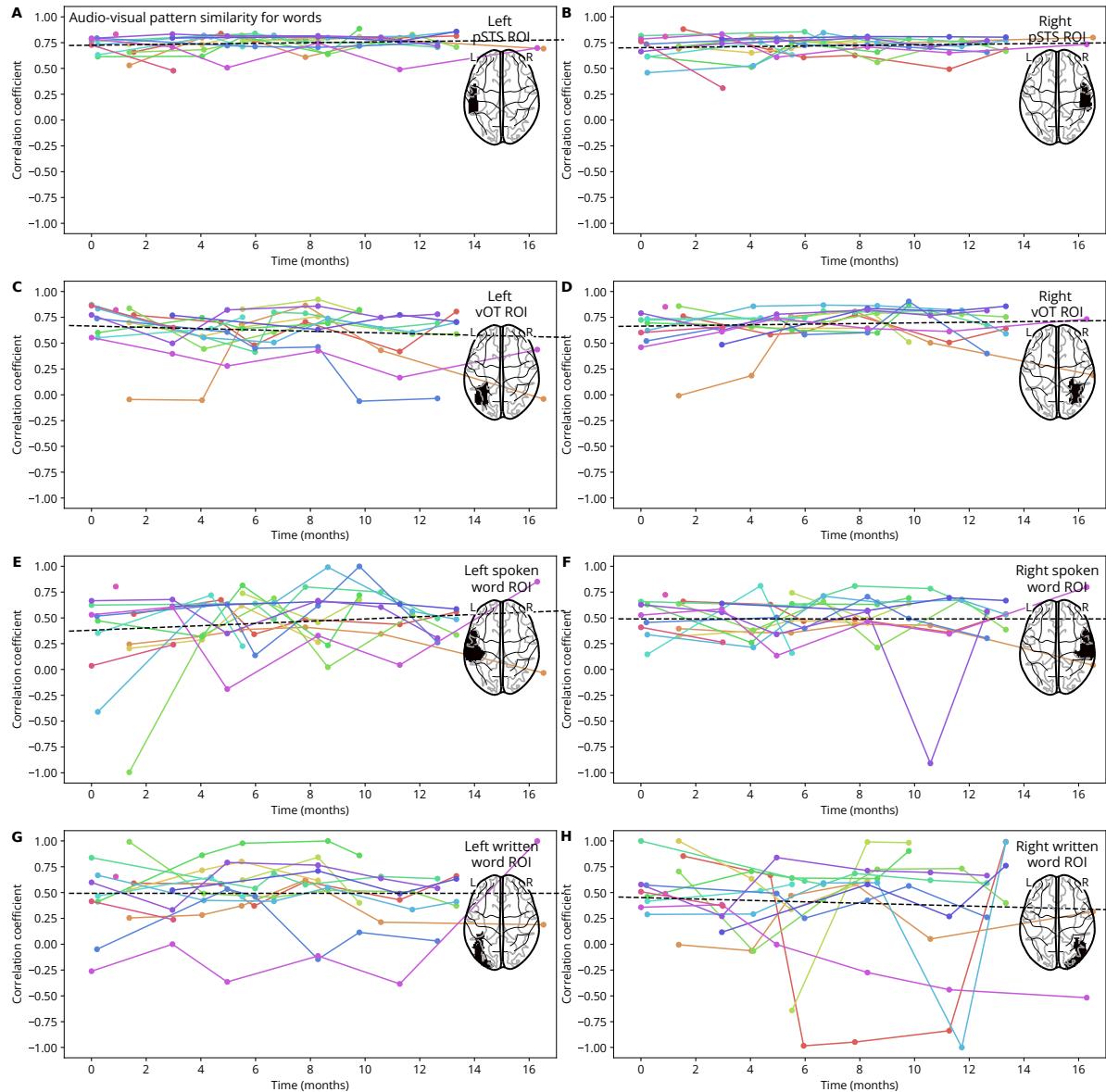


Figure 11. Audio-visual pattern similarity for words. Each panel A–H shows the development of audio-visual pattern similarity for the contrast of word blocks vs. fixation baseline over the course of the study in one region of interest (ROI). Audio-visual pattern similarity was computed by correlating the beta weights of all voxels inside the ROI for the auditory contrast (spoken words vs. fixation baseline) with those for the visual condition (written words vs. fixation baseline). ROIs were defined either anatomically (posterior superior temporal sulcus [pSTS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 5). Colored dots and lines indicate the pattern similarities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

ity from one block to the next, (c) no optimization of the order of blocks with regards to the efficiency of the experimental design (Henson, 2004), and (d) individual stimuli being shorter (1 s) than the TR (2.55 s), which precluded any stimulus-level analysis (e.g., with regards to BOLD pattern similarity and stability).

Implications for future research

Based on the limitations mentioned directly above, we advise the reader not to draw any strong conclusions from the empirical effects reported in this paper (or any lack thereof). However, we still believe that this study can

stimulate and facilitate future research in at least two important ways.

First, our overall study design can be seen as a proof of principle for overcoming geographic and cultural biases in developmental cognitive neuroscience. Most previous research on learning to read and its neural correlates has been carried out with participants from the Global North and in languages with alphabetic writing systems. In the future, it will be crucial to test if the empirical findings and theories derived from these populations will generalize to different cultures and writing system, or if these require novel hypotheses and theories (Frost, 2012; Share, 2008, 2014, 2021). We have shown that it is indeed possible—though certainly challenging—to collect

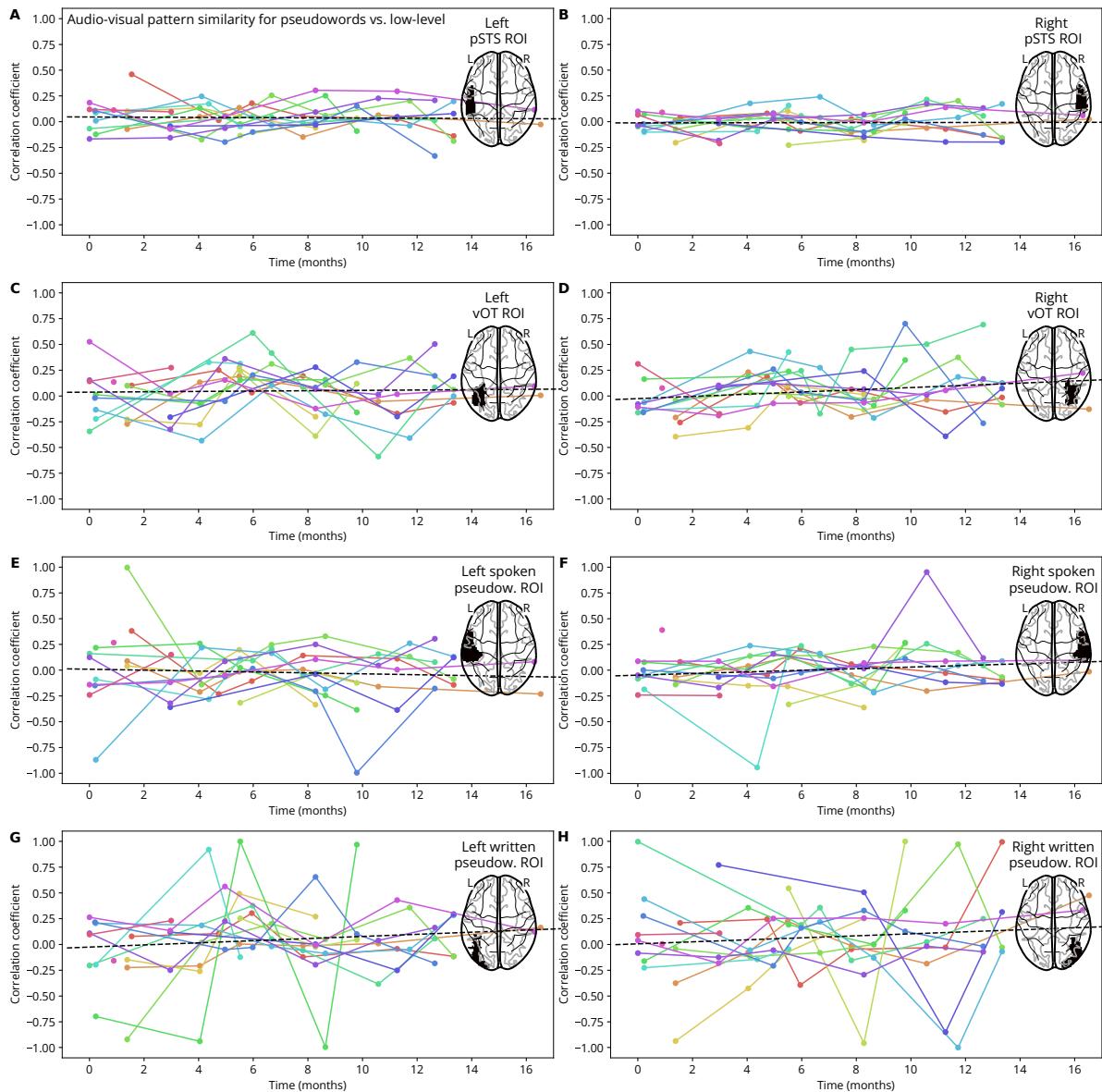


Figure 12. Audio-visual pattern similarity for pseudowords vs. low-level controls. Each panel A–H shows the development of audio-visual pattern similarity for the contrast of pseudoword blocks vs. low-level sensory control blocks over the course of the study in one region of interest (ROI). Audio-visual pattern similarity was computed by correlating the beta weights of all voxels inside the ROI for the auditory contrast (spoken pseudowords vs. noise-vocoded speech) with those for the visual condition (written pseudowords false fonts). ROIs were defined either anatomically (posterior superior temporal sulcus [STS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 6). Colored dots and lines indicate the pattern similarities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

relatively high-quality neuroimaging data in geographic areas and from socio-economic strata that have traditionally been neglected in cognitive neuroscience (Blasi et al., 2022; Henrich et al., 2010). In the context of learning to read and other learning interventions, such populations can even offer unique advantages, e.g., because illiteracy or lack of public schooling can allow for targeted and controlled learning interventions for limited periods of time, as we attempted in the present study.⁷ We therefore encourage more researchers to engage in cross-cultural collaboration projects, to test the generalizability of empirical findings and theories across different geographic and socio-economic settings, or at the very least to explicitly acknowledge the limited epistemic scope of findings

obtained solely from convenience samples in the Global North.

Second, we developed a novel computational approach for analyzing longitudinal fMRI data with whole-brain linear mixed-effects models. Compared to traditional statistical models for group-level fMRI analysis (typically variants of *t*-tests and analyses of variance [ANOVAs]), linear mixed-effects models come with a number of advantages:

⁷Note that in collaboration with a local non-governmental organization, we ensured that participants in our study would get enrolled at a public primary school after the conclusion of the study, something they would otherwise most likely never have gotten the opportunity to.

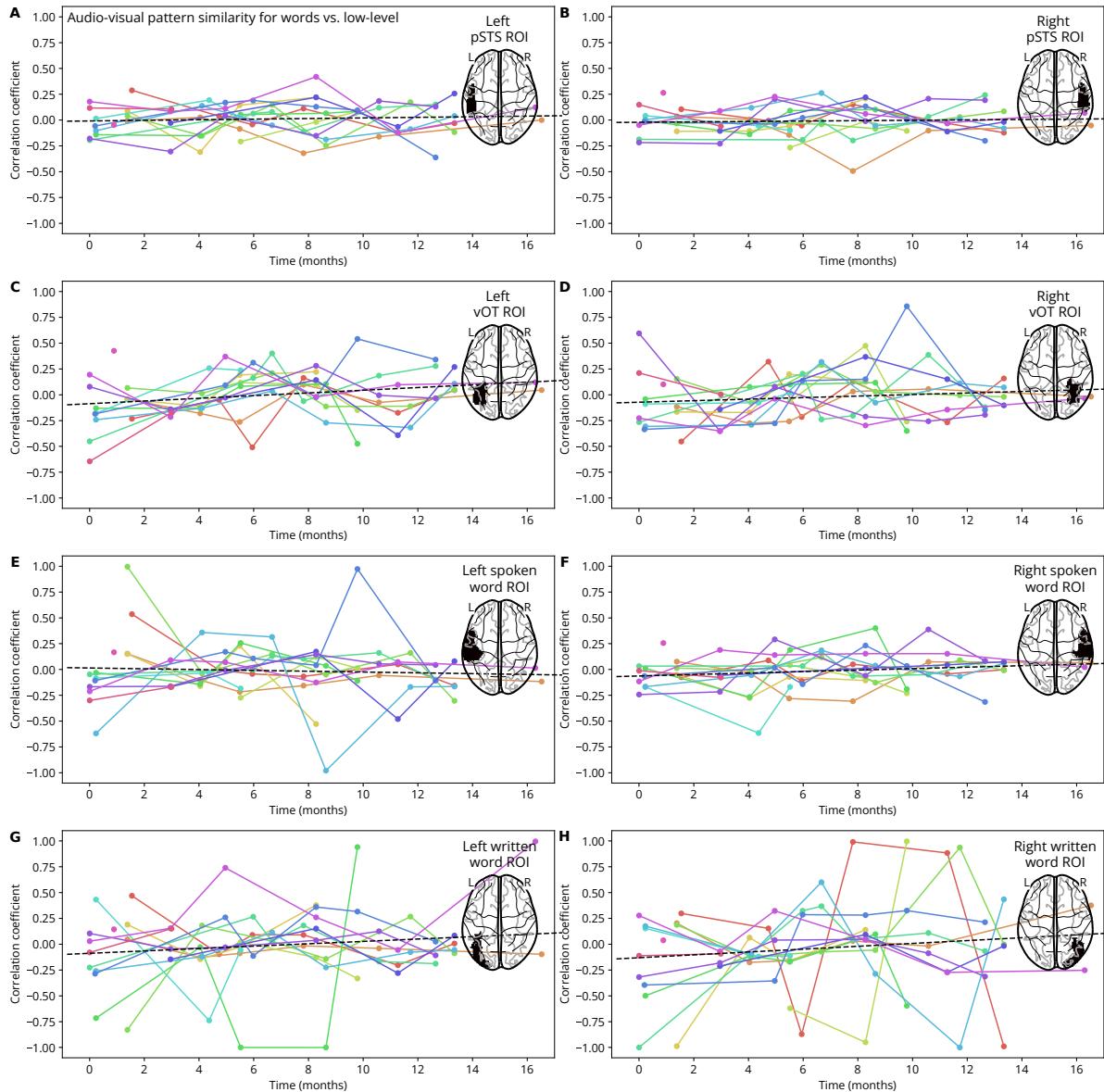


Figure 13. Audio-visual pattern similarity for words vs. low-level controls. Each panel A–H shows the development of audio-visual pattern similarity for the contrast of word blocks vs. low-level sensory control blocks over the course of the study in one region of interest (ROI). Audio-visual pattern similarity was computed by correlating the beta weights of all voxels inside the ROI for the auditory contrast (spoken words vs. noise-vocoded speech) with those for the visual condition (written words false fonts). ROIs were defined either anatomically (posterior superior temporal sulcus [STS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 7). Colored dots and lines indicate the pattern similarities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

- They can adequately control for multiple repeated measures obtained from the same participants.
- They can model longitudinal time as a continuous predictor variable instead of requiring discrete “sessions.”
- They do not require a balanced study design (e.g., the same number of sessions for all participants), and are therefore better able to handle dropout and missing data.
- They can be used to include multiple predictor variables of different type (e.g., linear time, quadratic time, group membership, gender).
- They can be used to relate within-individual fMRI effects to between-individual characteristics (i.e.,

brain-behavior associations; but note that this typically requires large sample sizes; Marek et al., 2022).

There have been previous implementations of whole-brain linear mixed-effects models, typically using the R programming language (Chen et al., 2013; Madhyastha et al., 2018). We believe that our implementation using the Julia programming language (Bates et al., 2024; Bezanson et al., 2017) provides another step forward towards their wider applicability, as model fitting in Julia is orders of magnitude faster and less prone to convergence errors, both of which is critical when testing the same model on hundreds of thousands of voxels in a mass-

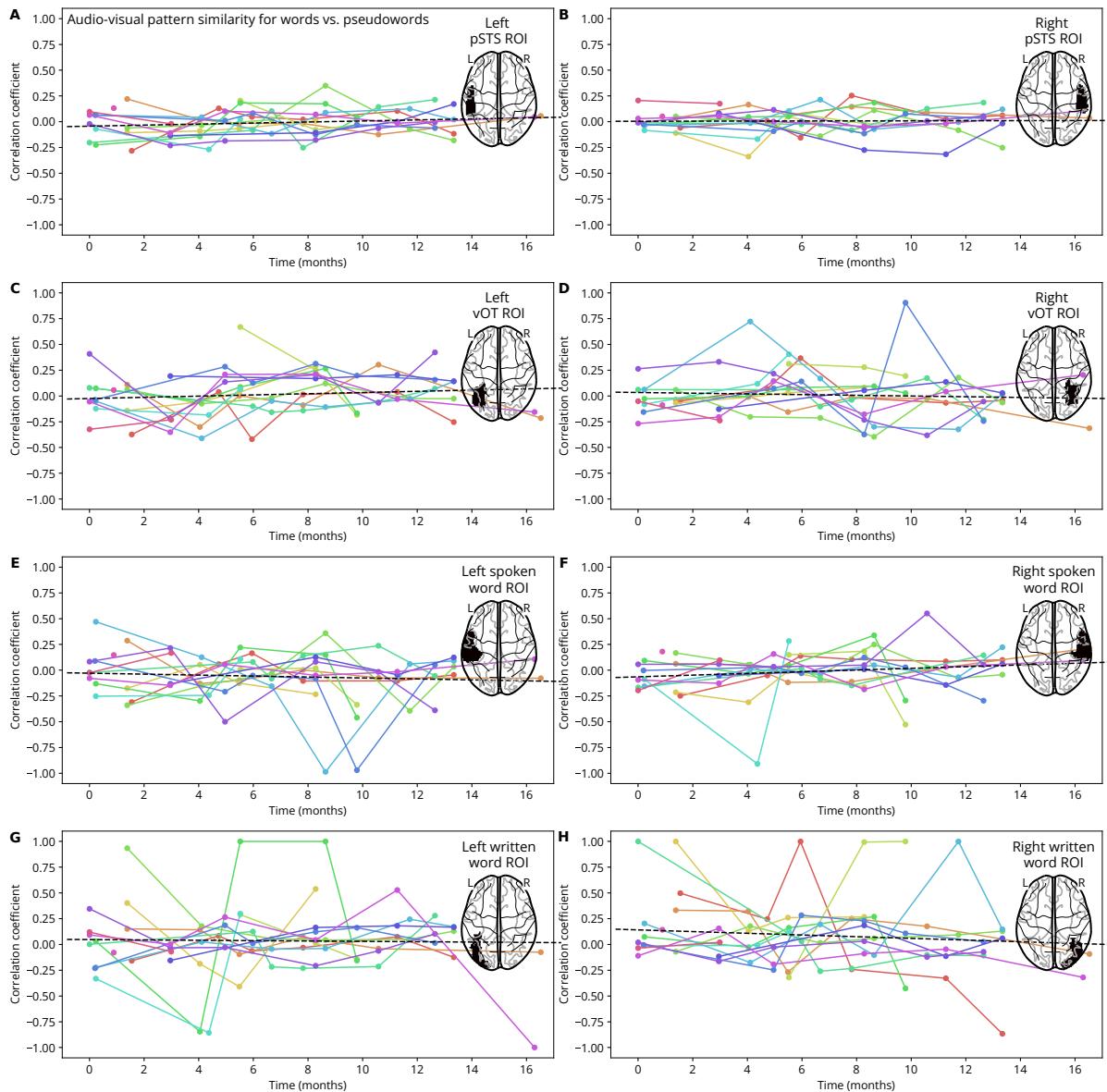


Figure 14. Audio-visual pattern similarity for words vs. pseudowords. Each panel A–H shows the development of audio-visual pattern similarity for the contrast of words blocks vs. pseudoword blocks over the course of the study in one region of interest (ROI). Audio-visual pattern similarity was computed by correlating the beta weights of all voxels inside the ROI for the auditory contrast (spoken words vs. spoken pseudowords) with those for the visual condition (written words vs. written pseudowords). ROIs were defined either anatomically (posterior superior temporal sulcus [STS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 8). Colored dots and lines indicate the pattern similarities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

univariate fashion. We combined this with a recently established method for whole-brain multiple comparison correction that is compatible with the outputs of the linear mixed-effects models to control the whole-brain family-wise error rate (Cox et al., 2017a, 2017b). We hope that this analysis approach and our code (available at <https://github.com/SkeideLab/SLANG-analysis/blob/SLANG/scripts/univariate.py>) can be reused for future longitudinal fMRI studies or other studies with complex data acquisition schedules that would benefit from the flexibility provided by linear mixed-effects models.

Conclusion

In this longitudinal fMRI study, we repeatedly measured children's brain activity in response to written and spoken words as they were learning to read. We found some evidence for non-linear changes in word-related BOLD activity amplitude but not in the direction or anatomical locations that we had hypothesized. We also found some evidence that learning to read leads to more similar BOLD activity patterns for visual and auditory stimuli in ventral occipito-temporal cortex. We found no evidence that learning to read leads to more stable BOLD activity response patterns in response to words vs. pseudowords. While this study had a number of limitations, including a

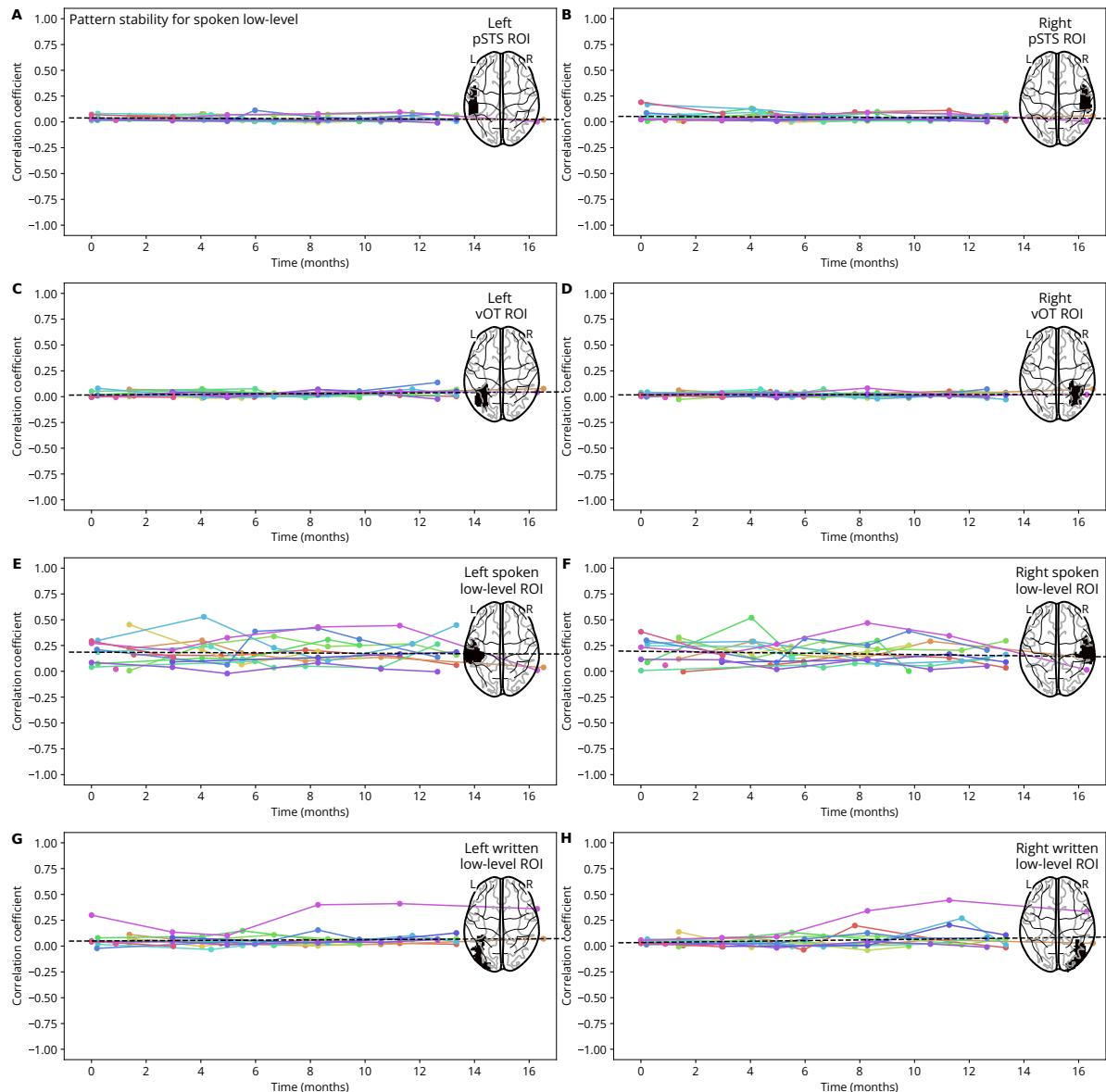


Figure 15. Pattern stability for spoken low-level controls. Each panel A–H shows the development of within-condition pattern stability for the contrast of spoken low-level sensory control blocks vs. fixation baseline over the course of the study in one region of interest (ROI). Pattern stability was computed by estimating beta weights separately for each spoken low-level sensory control block in the experiment, then computing all pairwise correlations of these beta weights inside the ROI, and then averaging all of these pairwise correlations to obtain a single correlation value for each participant and session. ROIs were defined either anatomically (posterior superior temporal sulcus [STS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 3). Colored dots and lines indicate the pattern stabilities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

very small sample size, future studies might want to reuse our overall study design, including the focus on a participant population and writing system typically neglected in developmental cognitive neuroscience, as well as our analysis methods for longitudinal whole-brain fMRI analysis.

References

- Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8, 1–10.
<https://doi.org/10.3389/fninf.2014.00014>
- Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, 12(1), 26–41.
<https://doi.org/10.1016/j.media.2007.06.004>

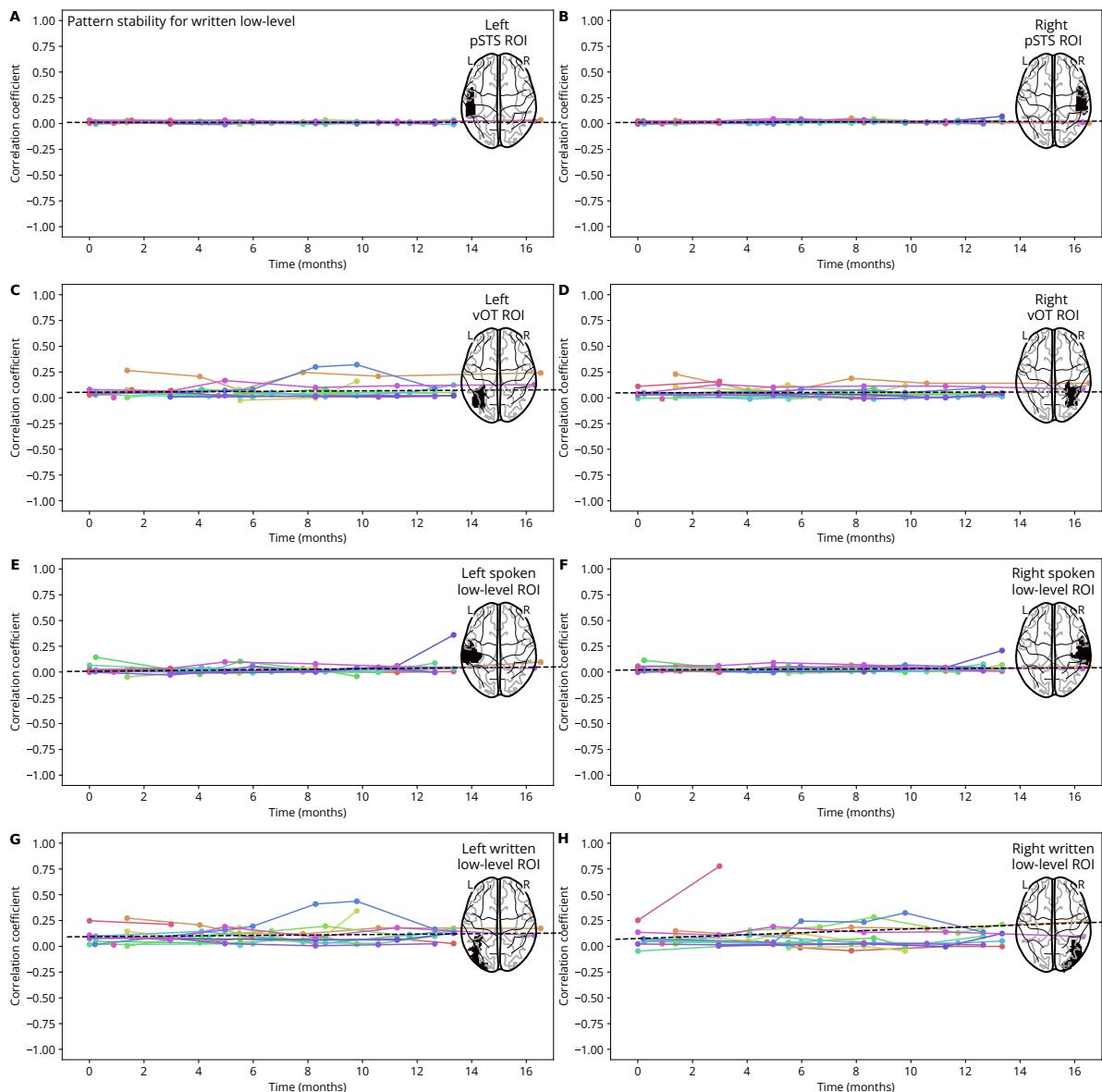


Figure 16. Pattern stability for written low-level controls. Each panel A–H shows the development of within-condition pattern stability for the contrast of written low-level sensory control blocks vs. fixation baseline over the course of the study in one region of interest (ROI). Pattern stability was computed by estimating beta weights separately for each written low-level sensory control block in the experiment, then computing all pairwise correlations of these beta weights inside the ROI, and then averaging all of these pairwise correlations to obtain a single correlation value for each participant and session. ROIs were defined either anatomically (posterior superior temporal sulcus [pSTS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 3). Colored dots and lines indicate the pattern stabilities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

- Bates, D., Alday, P., Kleinschmidt, D., Calderón, J. B. S., Zhan, L., Noack, A., Bouchet-Valat, M., Arslan, A., Kelman, T., Baldassari, A., Ehinger, B., Karrasch, D., Saba, E., Quinn, J., Hatherly, M., Piibelehto, M., Mogensen, P. K., Babayan, S., Holy, T., ... Nazarathy, Y. (2024). *JuliaStats/MixedModels.jl: V4.25.3*. Zenodo. <https://doi.org/10.5281/zenodo.13174525>
- Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1), 90–101. <https://doi.org/10.1016/j.neuroimage.2007.04.042>

- Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1), 65–98. <https://doi.org/10.1137/141000671>
- Blasi, D. E., Henrich, J., Adamou, E., Kemmerer, D., & Majid, A. (2022). Over-reliance on English hinders cognitive science. *Trends in Cognitive Sciences*, 26(12), 1153–1170. <https://doi.org/10.1016/j.tics.2022.09.015>
- Blau, V., Reithler, J., van Atteveldt, N., Seitz, J., Gerretsen, P., Goebel, R., & Blomert, L. (2010). Deviant processing of letters and speech sounds as proximate cause of reading failure: A functional magnetic resonance imaging study of dyslexic

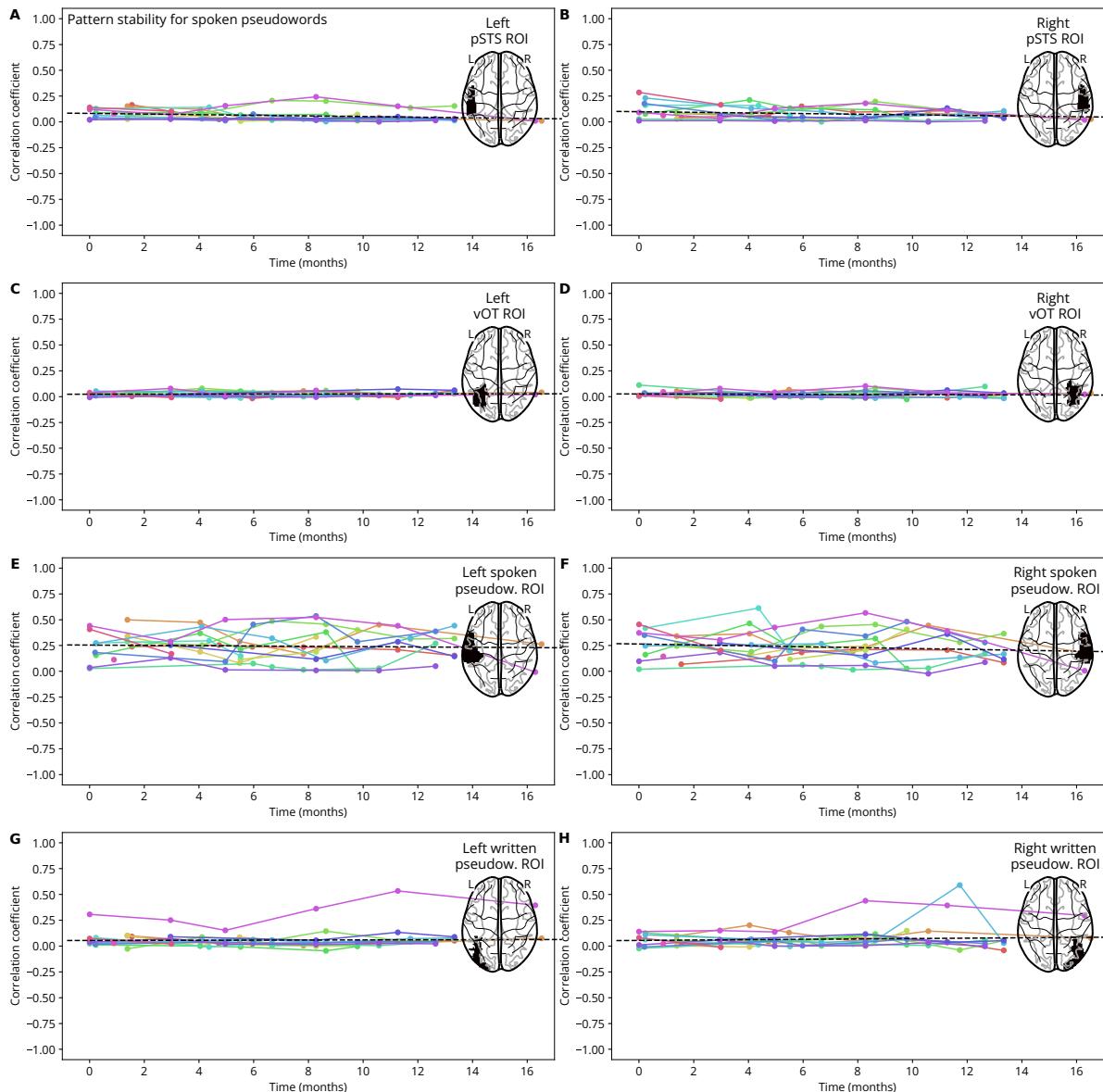


Figure 17. Pattern stability for spoken pseudowords. Each panel A–H shows the development of within-condition pattern stability for the contrast of spoken pseudoword blocks vs. fixation baseline over the course of the study in one region of interest (ROI). Pattern stability was computed by estimating beta weights separately for each spoken pseudoword block in the experiment, then computing all pairwise correlations of these beta weights inside the ROI, and then averaging all of these pairwise correlations to obtain a single correlation value for each participant and session. ROIs were defined either anatomically (posterior superior temporal sulcus [STS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 4). Colored dots and lines indicate the pattern stabilities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

children. *Brain*, 133(3), 868–879.
<https://doi.org/10.1093/brain/awz311>

Brem, S., Maurer, U., Kronbichler, M., Schurz, M., Richlan, F., Blau, V., Reithler, J., van der Mark, S., Schulz, E., Bucher, K., Moll, K., Landerl, K., Martin, E., Goebel, R., Schulte-Körne, G., Blomert, L., Wimmer, H., & Brandeis, D. (2020). Visual word form processing deficits driven by severity of reading impairments in children with developmental dyslexia. *Scientific Reports*, 10(1), 18728.
<https://doi.org/10.1038/s41598-020-15111-8>

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size

undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14(5), 365–376.
<https://doi.org/10.1038/nrn3475>

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W. R., Iversen, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science*, 276(5312), 593–596.
<https://doi.org/10.1126/science.276.5312.593>

Chen, G., Saad, Z. S., Britton, J. C., Pine, D. S., & Cox, R. W. (2013). Linear mixed-effects modeling approach to fMRI group analysis. *NeuroImage*, 73, 176–190.
<https://doi.org/10.1016/j.neuroimage.2013.01.047>

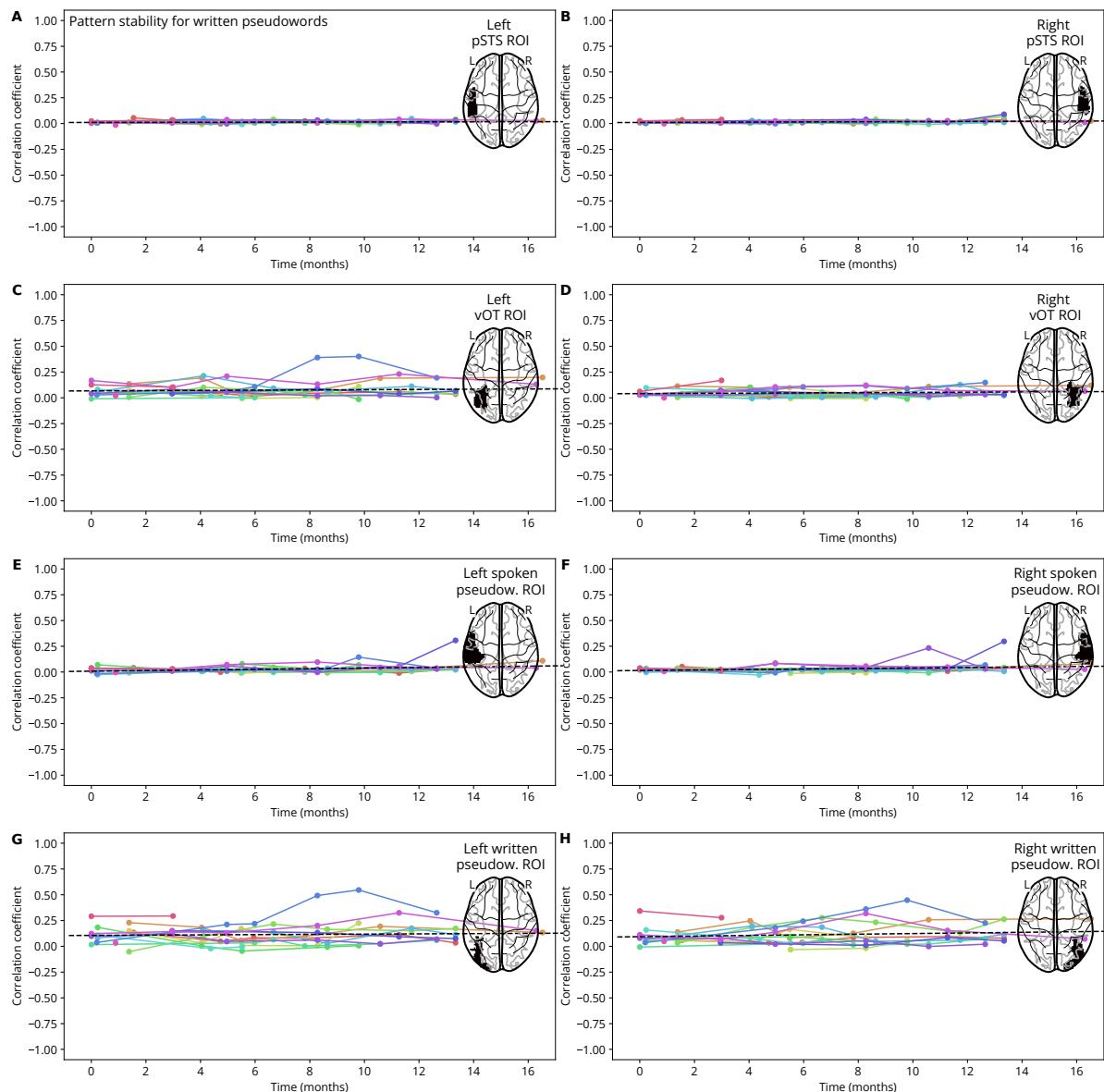


Figure 18. Pattern stability for written pseudowords. Each panel A–H shows the development of within-condition pattern stability for the contrast of written pseudoword blocks vs. fixation baseline over the course of the study in one region of interest (ROI). Pattern stability was computed by estimating beta weights separately for each written pseudoword block in the experiment, then computing all pairwise correlations of these beta weights inside the ROI, and then averaging all of these pairwise correlations to obtain a single correlation value for each participant and session. ROIs were defined either anatomically (posterior superior temporal sulcus [STS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 4). Colored dots and lines indicate the pattern stabilities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

Ciric, R., Thompson, W. H., Lorenz, R., Goncalves, M., MacNicol, E., Markiewicz, C. J., Halchenko, Y. O., Ghosh, S. S., Gorgolewski, K. J., Poldrack, R. A., & Esteban, O. (2022). TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models. *Nature Methods*, 19, 1568–1571.
<https://doi.org/10.1038/s41592-022-01681-2>

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M.-A., & Michel, F. (2000). The Visual Word Form Area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients. *Brain*, 123(2), 291–307.
<https://doi.org/10.1093/brain/123.2.291>

Coltheart, M. (2014). The neuronal recycling hypothesis for reading and the question of reading universals. *Mind & Language*, 29(3), 255–269.
<https://doi.org/10.1111/mila.12049>

Corsi, P. (1972). *Memory and the medial temporal region of the brain* [PhD thesis]. McGill University.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, 29(3), 162–173.
<https://doi.org/10.1006/cbmr.1996.0014>

Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017a). fMRI clustering and false-positive rates. *Proceedings of the National*

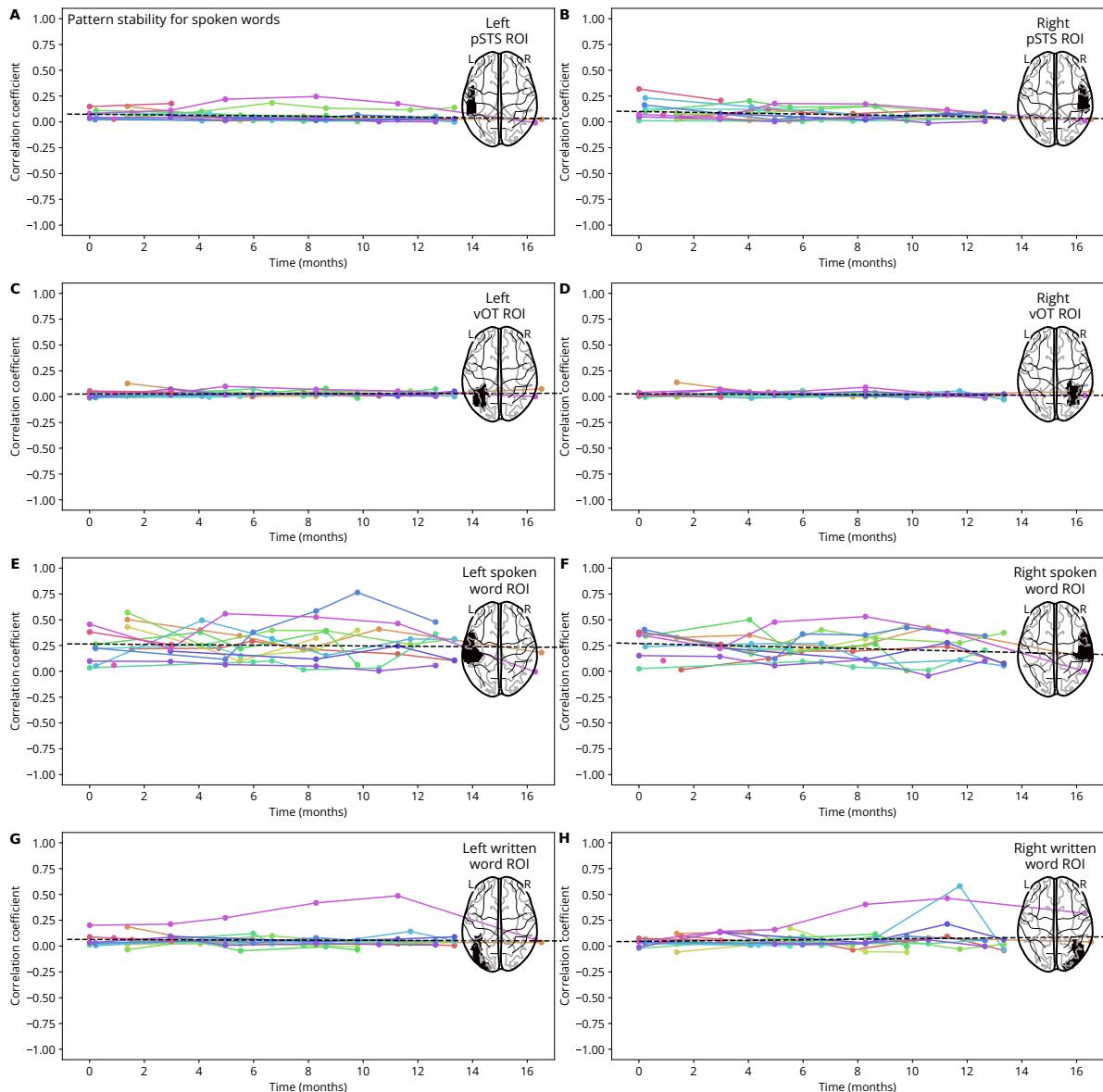


Figure 19. Pattern stability for spoken words. Each panel A–H shows the development of within-condition pattern stability for the contrast of spoken word blocks vs. fixation baseline over the course of the study in one region of interest (ROI). Pattern stability was computed by estimating beta weights separately for each spoken word block in the experiment, then computing all pairwise correlations of these beta weights inside the ROI, and then averaging all of these pairwise correlations to obtain a single correlation value for each participant and session. ROIs were defined either anatomically (posterior superior temporal sulcus [STS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 5). Colored dots and lines indicate the pattern stabilities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

- Academy of Sciences of the United States of America, 114(17), E3370–E3371.
<https://doi.org/10.1073/pnas.1614961114>
- Cox, R. W., Chen, G., Glen, D. R., Reynolds, R. C., & Taylor, P. A. (2017b). fMRI clustering in AFNI: False-positive rates redux. *Brain Connectivity*, 7(3), 152–171. <https://doi.org/10.1089/brain.2016.0475>
- Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. *NeuroImage*, 9(2), 179–194. <https://doi.org/10.1006/nimg.1998.0395>
- Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, 56(2), 384–398. <https://doi.org/10.1016/j.neuron.2007.10.004>

- Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences*, 15(6), 254–262. <https://doi.org/10.1016/j.tics.2011.04.003>
- Dehaene, S., Cohen, L., Morais, J., & Kolinsky, R. (2015). Illiterate to literate: Behavioural and cerebral changes induced by reading acquisition. *Nature Reviews Neuroscience*, 16(4), 234–244. <https://doi.org/10.1038/nrn3924>
- Dehaene, S., Le Clec'H, G., Poline, J.-B., Le Bihan, D., & Cohen, L. (2002). The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *NeuroReport*, 13(3), 321–325. <https://doi.org/10.1097/00001756-200203040-00015>

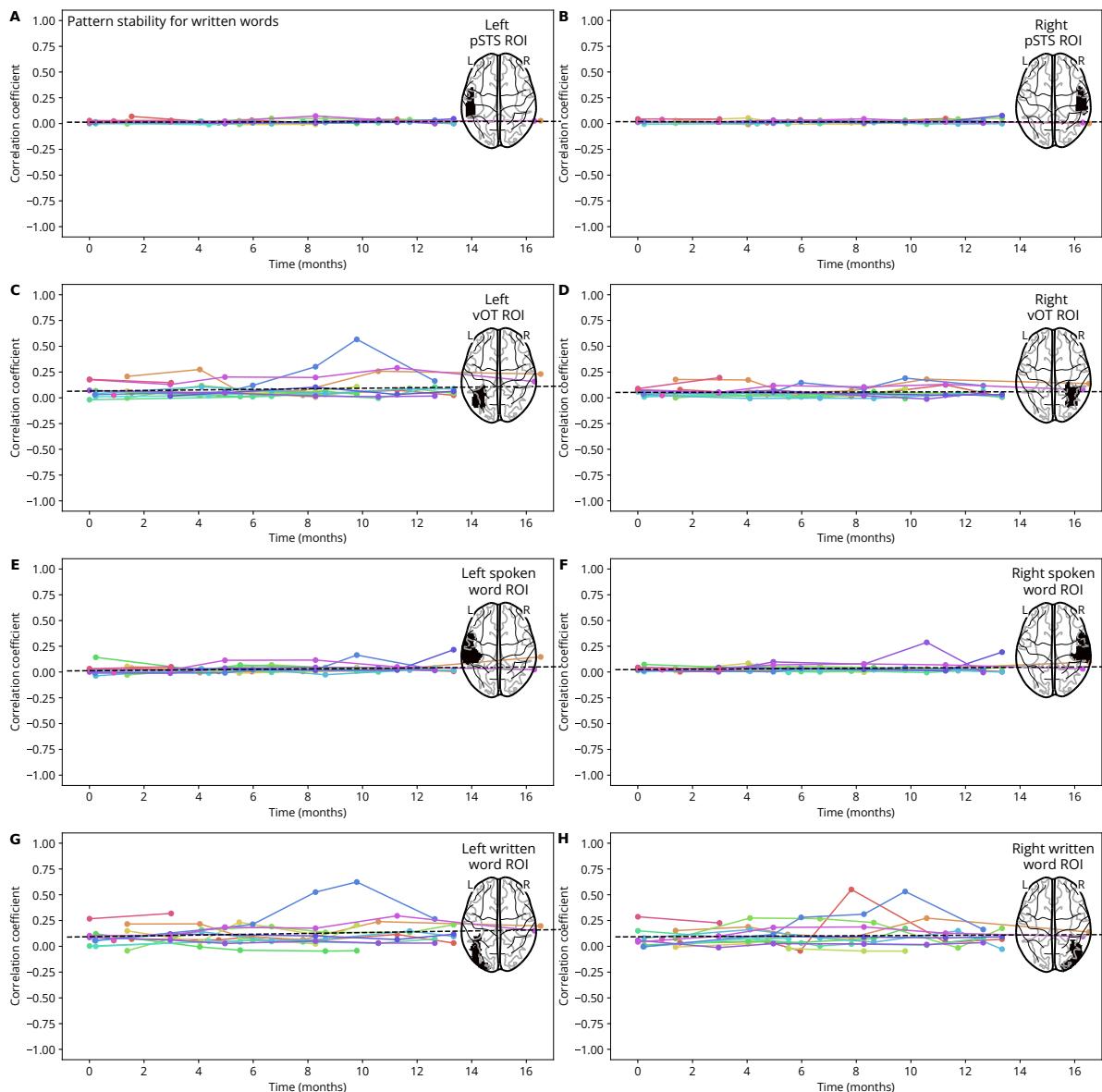


Figure 20. Pattern stability for written words. Each panel A–H shows the development of within-condition pattern stability for the contrast of written word blocks vs. fixation baseline over the course of the study in one region of interest (ROI). Pattern stability was computed by estimating beta weights separately for each written word block in the experiment, then computing all pairwise correlations of these beta weights inside the ROI, and then averaging all of these pairwise correlations to obtain a single correlation value for each participant and session. ROIs were defined either anatomically (posterior superior temporal sulcus [STS] and ventral occipito-temporal cortex [vOT]) or functionally from the significant clusters observed for the relevant baseline contrasts in the BOLD activity amplitude analysis (see Figure 5). Colored dots and lines indicate the pattern stabilities of individual participants over time. The black dashed line indicates the fit of a linear mixed-effects model to these data.

Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology*, 16(3), 1–34. <https://doi.org/10.1371/journal.pbio.2004103>

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, 9(2), 167–188. https://doi.org/10.1207/s15410704ssr0902_01

Embleton, K. V., Haroon, H. A., Morris, D. M., Ralph, M. A. L., & Parker, G. J. M. (2010). Distortion correction for diffusion-weighted MRI tractography and fMRI in the temporal lobes. *Human Brain Mapping*, 31(10), 1570–1587. <https://doi.org/10.1002/hbm.20959>

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, M., James D. andGoncalves, DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIprep. Zenodo. <https://doi.org/10.5281/zenodo.852659>

Esteban, O., Markiewicz, C., Blair, R. W., Moodie, C., Isik, A. I., Erramuzpe Aliaga, A., Kent, J., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R., & Gorgolewski, K. J. (2019). fMRIprep: A robust

- preprocessing pipeline for functional MRI. *Nature Methods*, 16, 111–116.
<https://doi.org/10.1038/s41592-018-0235-4>
- Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47, Supplement 1, S102. [https://doi.org/10.1016/S1053-8119\(09\)70884-5](https://doi.org/10.1016/S1053-8119(09)70884-5)
- Forseth, K. J., Kadipasaoglu, C. M., Conner, C. R., Hickok, G., Knight, R. T., & Tandon, N. (2018). A lexical semantic hub for heteromodal naming in middle fusiform gyrus. *Brain*, 141(7), 2112–2126. <https://doi.org/10.1093/brain/awy120>
- Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia*, 36(1), 67–81. <https://doi.org/10.1007/BF02648022>
- Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences*, 35(5), 263–279. <https://doi.org/10.1017/S0140525X11001841>
- Glasser, M. F., Coalson, T. S., Robinson, E. C., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615), 171–178. <https://doi.org/10.1038/nature18933>
- Gorgolewski, K. J., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, 5, 13. <https://doi.org/10.3389/fninf.2011.00013>
- Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F., ... Ghosh, S. (2018). Nipype. *Zenodo*. <https://doi.org/10.5281/zenodo.596855>
- Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1), 63–72. <https://doi.org/10.1016/j.neuroimage.2009.06.060>
- Hasenäcker, J., Schröter, P., & Schroeder, S. (2017). Investigating developmental trajectories of morphemes as reading units in German. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(7), 1093–1108. <https://doi.org/10.1037/xlm0000353>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Henson, R. (2004). Analysis of fMRI timeseries: Linear time-invariant models, event-related fMRI and optimal experimental design. In R. S. J. Frackowiak (Ed.), *Human Brain Function* (2nd ed., pp. 193–210). Academic Press.
- Hillis, A. E., Newhart, M., Heidler, J., Barker, P., Herskovits, E., & Degaonkar, M. (2005). The roles of the "visual word form area" in reading. *NeuroImage*, 24(2), 548–559. <https://doi.org/10.1016/j.neuroimage.2004.08.026>
- Hirshorn, E. A., Li, Y., Ward, M. J., Richardson, R. M., Fiez, J. A., & Ghuman, A. S. (2016). Decoding and disrupting left midfusiform gyrus activity during word reading. *Proceedings of the National Academy of Sciences*, 113(29), 8162–8167. <https://doi.org/10.1073/pnas.1604126113>
- Houston, S. D. (Ed.). (2004). *The first writing: Script invention as history and process*. Cambridge University Press.
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), 825–841. <https://doi.org/10.1006/nimg.2002.1132>
- Kessels, R. P., van Zandvoort, M. J., Postma, A., Kappelle, L. J., & de Haan, E. H. (2000). The Corsi Block-Tapping Task: Standardization and normative data. *Applied Neuropsychology*, 7(4), 252–258. https://doi.org/10.1207/S15324826AN0704_8
- Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, 13(2), e1005350. <https://doi.org/10.1371/journal.pcbi.1005350>
- Kubota, E., Grill-Spector, K., & Nordt, M. (2023). Rethinking cortical recycling in ventral temporal cortex. *Trends in Cognitive Sciences*. <https://doi.org/10.1016/j.tics.2023.09.006>
- Liu, T. T. (2016). Noise contributions to the fMRI signal: An overview. *NeuroImage*, 143, 141–151. <https://doi.org/10.1016/j.neuroimage.2016.09.008>
- Madhyastha, T., Peverill, M., Koh, N., McCabe, C., Flournoy, J., Mills, K., King, K., Pfeifer, J., & McLaughlin, K. A. (2018). Current methods and limitations for longitudinal fMRI analysis across development. *Developmental Cognitive Neuroscience*, 33, 118–128. <https://doi.org/gdz7rh>
- Marek, S., Tervo-Clemmens, B., Calabro, F. J., Montez, D. F., Kay, B. P., Hatoum, A. S., Donohue, M. R., Foran, W., Miller, R. L., Hendrickson, T. J., Malone, S. M., Kandala, S., Feczkó, E., Miranda-Dominguez, O., Graham, A. M., Earl, E. A., Perrone, A. J., Cordova, M., Doyle, O., ... Dosenbach, N. U. F. (2022). Reproducible brain-wide association studies require thousands of individuals. *Nature*, 603(7902), 654–660. <https://doi.org/10.1038/s41586-022-04492-9>
- McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences*, 7(7), 293–299. [https://doi.org/10.1016/S1364-6613\(03\)00134-7](https://doi.org/10.1016/S1364-6613(03)00134-7)
- Melby-Lervåg, M., Lyster, S.-A. H., & Hulme, C. (2012). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin*, 138(2), 322–352. <https://doi.org/10.1037/a0026744>
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for

- processing information. *Psychological Review*, 63(2), 81–97. <https://doi.org/10.1037/h0043158>
- Murphy, K. A., Jogia, J., & Talcott, J. B. (2019). On the neural basis of word reading: A meta-analysis of fMRI evidence using activation likelihood estimation. *Journal of Neurolinguistics*, 49, 71–83. <https://doi.org/10.1016/j.jneuroling.2018.08.005>
- Nordt, M., Gomez, J., Natu, V. S., Rezai, A. A., Finzi, D., Kular, H., & Grill-Spector, K. (2021). Cortical recycling in high-level visual cortex during childhood development. *Nature Human Behaviour*, 1–12. <https://doi.org/10.1038/s41562-021-01141-5>
- Patriat, R., Reynolds, R. C., & Birn, R. M. (2017). An improved model of motion-related signal changes in fMRI. *NeuroImage*, 144, Part A, 74–82. <https://doi.org/10.1016/j.neuroimage.2016.08.051>
- Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, 84, 320–341. <https://doi.org/10.1016/j.neuroimage.2013.08.048>
- Price, C. J., & Devlin, J. T. (2003). The myth of the visual word form area. *NeuroImage*, 19(3), 473–481. [https://doi.org/10.1016/S1053-8119\(03\)00084-3](https://doi.org/10.1016/S1053-8119(03)00084-3)
- Rao, C., Midha, R., Midya, V., Sumathi, T. A., Singh, N., Oberoi, G., Kar, B., Currawala, K., Khan, M., Rao, P., Shukla, S., & Vaidya, K. (2015). *Dyslexia Assessment for Languages of India (DALI)*. National Brain Research Centre. <https://doi.org/10.13140/RG.2.2.14696.32005>
- Rao, C., Sumathi, T. A., Midha, R., Oberoi, G., Kar, B., Khan, M., Vaidya, K., Midya, V., Raman, N., Gajre, M., & Singh, N. C. (2021). Development and standardization of the DALI-DAB (dyslexia assessment for languages of India - dyslexia assessment battery). *Annals of Dyslexia*, 71(3), 439–457. <https://doi.org/10.1007/s11881-021-00227-z>
- Raven, J., & Raven, J. (2003). Raven progressive matrices. In R. S. McCallum (Ed.), *Handbook of Nonverbal Assessment* (pp. 223–237). Springer US. https://doi.org/10.1007/978-1-4615-0153-4_11
- Reuter, M., Rosas, H. D., & Fischl, B. (2010). Highly accurate inverse consistent registration: A robust approach. *NeuroImage*, 53(4), 1181–1196. <https://doi.org/10.1016/j.neuroimage.2010.07.020>
- Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64(1), 240–256. <https://doi.org/10.1016/j.neuroimage.2012.08.052>
- Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an “outlier” orthography. *Psychological Bulletin*, 134(4), 584–615. <https://doi.org/10.1037/0033-2909.134.4.584>
- Share, D. L. (2014). Alphabetism in reading science. *Frontiers in Psychology*, 5, 752. <https://doi.org/10.3389/fpsyg.2014.00752>
- Share, D. L. (2021). Is the science of reading just the science of reading English? *Reading Research Quarterly*, 56(S1), 391–402. <https://doi.org/10.1002/rrq.401>
- Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews Neuroscience*, 17(5), 323–332. <https://doi.org/10.1038/nrn.2016.23>
- Szucs, D., & Ioannidis, J. P. A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLOS Biology*, 15(3), e2000797. <https://doi.org/10.1371/journal.pbio.2000797>
- Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6), 1310–1320. <https://doi.org/10.1109/TMI.2010.2046908>
- van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, 43(2), 271–282. <https://doi.org/10.1016/j.neuron.2004.06.025>
- Van Rossum, G., & Drake, F. L. (2009). *Python 3 reference manual*. CreateSpace.
- Wenger, E., Brozzoli, C., Lindenberger, U., & Lövdén, M. (2017). Expansion and renormalization of human brain structure during skill acquisition. *Trends in Cognitive Sciences*, 21(12), 930–939. <https://doi.org/10.1016/j.tics.2017.09.008>
- Wilkinson, G. S., & Robertson, G. J. (2017). *Wide range achievement test, fifth edition – India (WRAT5 – INDIA)*. Pearson Clinical.
- Wilson, S. M., Bautista, A., & McCarron, A. (2018). Convergence of spoken and written language processing in the superior temporal sulcus. *NeuroImage*, 171, 62–74. <https://doi.org/gc8hq4>
- Woolnough, O., Donos, C., Curtis, A., Rollo, P. S., Roccaforte, Z. J., Dehaene, S., Fischer-Baum, S., & Tandon, N. (2022). A spatiotemporal map of reading aloud. *Journal of Neuroscience*, 42(27), 5438–5450. <https://doi.org/10.1523/JNEUROSCI.2324-21.2022>
- Woolnough, O., Donos, C., Rollo, P. S., Forseth, K. J., Lakretz, Y., Crone, N. E., Fischer-Baum, S., Dehaene, S., & Tandon, N. (2020). Spatiotemporal dynamics of orthographic and lexical processing in the ventral visual pathway. *Nature Human Behaviour*, 1–10. <https://doi.org/10.1038/s41562-020-00982-w>
- Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1), 45–57. <https://doi.org/10.1109/42.906424>

Instant Effects of Semantic Information on Visual Perception

✉ Alexander Enge,^{1,2} Franziska Süß,³ and Rasha Abdel Rahman^{1,4}

¹Department of Psychology, Humboldt-Universität zu Berlin, 12489 Berlin, Germany, ²Max Planck Institute for Human Cognitive, Research Group Learning in Early Childhood, and Brain Sciences, 04103, Leipzig, Germany, ³Fachhochschule des Mittelstands, 96050, Bamberg, Germany, and

⁴Cluster of Excellence “Science of Intelligence,” 10587, Berlin, Germany

Does our perception of an object change once we discover what function it serves? We showed human participants ($n = 48$, 31 females and 17 males) pictures of unfamiliar objects either together with keywords matching their function, leading to semantically informed perception, or together with nonmatching keywords, resulting in uninformed perception. We measured event-related potentials to investigate at which stages in the visual processing hierarchy these two types of object perception differed from one another. We found that semantically informed compared with uninformed perception was associated with larger amplitudes in the N170 component (150–200 ms), reduced amplitudes in the N400 component (400–700 ms), and a late decrease in alpha/beta band power. When the same objects were presented once more without any information, the N400 and event-related power effects persisted, and we also observed enlarged amplitudes in the P1 component (100–150 ms) in response to objects for which semantically informed perception had taken place. Consistent with previous work, this suggests that obtaining semantic information about previously unfamiliar objects alters aspects of their lower-level visual perception (P1 component), higher-level visual perception (N170 component), and semantic processing (N400 component, event-related power). Our study is the first to show that such effects occur instantly after semantic information has been provided for the first time, without requiring extensive learning.

Key words: event-related potentials; objects; semantic knowledge; visual perception

Significance Statement

There has been a long-standing debate about whether or not higher-level cognitive capacities, such as semantic knowledge, can influence lower-level perceptual processing in a top-down fashion. Here we could show, for the first time, that information about the function of previously unfamiliar objects immediately influences cortical processing within less than 200 ms. Of note, this influence does not require training or experience with the objects and related semantic information. Therefore, our study is the first to show effects of cognition on perception while ruling out the possibility that prior knowledge merely acts by preactivating or altering stored visual representations. Instead, this knowledge seems to alter perception online, thus providing a compelling case against the impenetrability of perception by cognition.

Introduction

Does our perception of an object change once we discover what function it serves? This question speaks to the long-standing debate about the cognitive (im-)penetrability of perception by

higher-level capacities, such as semantic knowledge or language. According to one view, these cognitive capacities kick in only after the retinal input has been processed by a specialized module for visual perception (Fodor, 1983). This module is supposed to be encapsulated from higher-level inputs and processes visual information in a feedforward fashion, progressing from lower to higher areas representing increasingly complex shapes and, eventually, whole objects (DiCarlo et al., 2012). This cognitive impenetrability hypothesis is challenged by another view, namely, predictive coding theories that posit that higher-level areas influence ongoing perceptual processing early on by sending predictions down to lower-level areas (Churchland et al., 1994; Ahissar and Hochstein, 2004; Yuille and Kersten, 2006; Friston and Kiebel, 2009; Clark, 2013; Lupyan, 2015; Thierry, 2016; Teufel and Nanay, 2017; Lupyan et al., 2020). This view is supported by differences in ratings,

Received Oct. 25, 2022; revised Mar. 16, 2023; accepted Apr. 17, 2023.

Author contributions: A.E., F.S., and R.A.R. designed research; A.E. performed research; A.E. analyzed data; A.E. wrote the first draft of the paper; F.S. and R.A.R. edited the paper.

This work was supported by Deutsche Forschungsgemeinschaft, Germany's Excellence Strategy (EXC 2002/1 "Science of Intelligence") Project 390523135; and Grants AB277-5 and AB277-6 to R.A.R. We thank Nele Langosch for assistance with the preparation of materials and data collection; Olaf Dimigen and the Abdel Rahman laboratory for Neurocognitive Psychology for insightful discussions; Guido Kiecker for technical support; and two reviewers for comments on a previous version of the paper.

The authors declare no competing financial interests.

Correspondence should be addressed to Alexander Enge at alexander.enge@hu-berlin.de.

<https://doi.org/10.1523/JNEUROSCI.2038-22.2023>

Copyright © 2023 the authors

detection rates, or reaction times for visual stimuli depending on their emotional (e.g., Phelps et al., 2006), linguistic (e.g., Slivac et al., 2021), or semantic (e.g., Gauthier et al., 2003) content. However, some of these studies received legitimate criticism, for example, because comparisons between critical conditions included the confound of additionally comparing different visual stimuli, or for not being able to distinguish between perceptual or postperceptual loci of tentative top-down effects based on behavioral measures (Firestone and Scholl, 2016).

Here we use event-related potentials (ERPs) measured from the human EEG to mitigate most of these concerns: The temporal resolution of ERPs makes it possible to probe how early influences of high-level (e.g., semantic) information can be detected (Athanasopoulos and Casaponsa, 2020). When comparing objects after learning different amounts of semantic information about them, previous studies by our laboratory and others revealed differences in late ERP components associated with semantic processing (N400 component) and also in the visual P1 component (Abdel Rahman and Sommer, 2008; Samaha et al., 2018; Maier and Abdel Rahman, 2019; Weller et al., 2019). The early peak of the P1 and its source in the occipital cortex (Abdel Rahman and Sommer, 2008) point to an early effect of semantic knowledge on visual perception. It is less clear, however, if this effect acts in an online fashion (i.e., directly modulating perceptual processing, in accordance with predictive coding theories), or more indirectly, by altering stored visual object representations over the course of learning, which would then get reactivated once the object is re-encountered later on (i.e., reflecting offline differences in visual object representations or prototypes that have been built up over the course of learning) (Palmeri and Tarr, 2008).

To answer this question, we measured ERPs in response to unfamiliar objects directly while participants gained a semantic understanding of their function. We presented half of the objects with matching keywords, allowing participants to understand what kind of object they were viewing, and the other half with nonmatching keywords, preventing participants from understanding what kind of object they were viewing. We then presented the same objects again to test for downstream effects of semantic information. We examined the influence of semantic information on ERPs associated with lower-level visual perception (P1 component), higher-level visual perception (N170 component), and semantic processing (N400 component). We hypothesized that semantic information would have instant effects on visual perception, by which we mean (1) that these effects can be observed on the same trial as the semantic information is being presented for the first time, and (2) that these effects are found not only in later, higher-level cognition-related ERP components (N400), but also in earlier, perception-related ERP components (P1 and/or N170), which would speak for predictive coding theories. We furthermore conducted an exploratory time-frequency analysis to test for effects on event-related power which, unlike ERP effects, do not need to be tightly phase-locked to stimulus onset.

Materials and Methods

Participants. Participants were 48 German native speakers (31 females and 17 males) with a mean age of 23.5 years (range 18–32 years) and no history of psychological disorder or treatment. No *a priori* power analysis was conducted. All participants were right-handed according to the Edinburgh Inventory (Oldfield, 1971) and reported normal or corrected-to-normal vision. They provided written informed consent before

starting the experiment and received a monetary compensation of €8 per hour for participating.

Stimuli. Stimuli consisted of 240 grayscale photographs of real-world objects. Of these, 120 stimuli were well-known everyday objects (e.g., a bicycle, a toothbrush). These served as filler stimuli of no interest. The other 120 stimuli were rare objects presumed to be unfamiliar to most participants (e.g., a galvanometer, a udu drum; see Online Table 1 at <https://doi.org/10.17605/osf.io/ulksb>). All stimuli were presented on a light blue background with a size of 207 × 207 pixels on a 19-inch LCD monitor with a resolution of 1280 × 1024 pixels and a refresh rate of 75 Hz. At a standardized viewing distance of 90 cm, the images subtended ∼3.9 degrees of participants' horizontal and vertical visual angle.

For each unfamiliar object, we created a pair of German keywords (a noun and a verb), describing the typical function or use of the object in a way that could be related to its visual features and their configuration (e.g., *Stromstärke, messen* [electric current, measuring]; *Tonpott, trommeln* [clay pot, drumming]; see Online Table 1). As our central experimental manipulation, half of the objects were presented together with keywords that matched their respective function, whereas the other half of the objects were presented together with nonmatching keywords (which would have matched a different object). The matching keywords were expected to induce semantically informed perception, that is, participants suddenly understanding what kind of object they were viewing. The nonmatching keywords were expected to prevent such an understanding and keep the perception of the object semantically uninformed. All participants saw each unfamiliar object with only one type of keywords (matching or nonmatching). This assignment of keywords to objects was counterbalanced across participants so that each object was presented with matching keywords (leading to semantically informed perception) and nonmatching keywords (leading to uninformed perception) to an equal number of participants. The experiment was programmed and displayed using Presentation software (Neurobehavioral Systems; www.neurobs.com).

As a manipulation check, we ran an online rating study where we presented 10 German speakers (3 female, 7 male, mean age = 25.3 years, range 20–35 years; none took part in the main study) with all 240 visual objects in random order and asked them to generate their own keywords that would describe the presumed function of the object. We used latent semantic analysis (Günther et al., 2019) with a word2vec embedding (deepset GmbH, Berlin, Germany; www.deepset.ai/german-word-embeddings) pretrained on the German Wikipedia to estimate semantic distances between these participant-generated keywords and the keywords used in our main EEG experiment (see Online Figs. 1 and 2). In brief, it was substantially easier for participants to come with correct descriptions of familiar objects ($\text{mean cosine distance } \pm \text{SE} = 0.78 \pm 0.01$) than for unfamiliar objects ($0.66 \pm 0.01, t_{(13)} = 7.21, p < 0.001$, linear mixed-effects model). Indeed, this similarity between participant-generated keywords for the unfamiliar objects and the object-matching keywords presented in the main experiment was only slightly higher than the similarity between participant-generated keywords and the object-nonmatching keywords presented in the main experiment ($0.64 \pm 0.01, t_{(24)} = 5.09, p < 0.001$, linear mixed-effects model), indicating that it was difficult for participants to know or guess the correct function of the unfamiliar objects.

Experimental design. The main EEG experiment consisted of three phases (see Fig. 1A). In the pre-insight phase, after written informed consent had been obtained and the EEG had been prepared, all 240 familiar and unfamiliar objects were presented once in random order and without any keywords. Each trial consisted of a fixation cross presented in the middle of the screen for 0.5 s, followed by the presentation of the object until participants made a response or until a timeout after 3 s. The intertrial interval was 0.5 s, and participants took a self-timed break after each block of 60 objects. The task, which was kept the same across all three phases of the experiment, was to classify each object using one of four response alternatives: (a) *Ich weiß, was das ist, oder habe eine starke Vermutung* [I know what this is or have a strong assumption], (b) *Ich habe eher eine Vermutung, was das ist* [I rather have an assumption what this is], (c) *Ich habe eher keine Vermutung, was das ist* [I rather have no assumption what this is], or (d) *Ich weiß nicht, was das ist, und habe auch keine Vermutung* [I don't know what this is and have no assumption].

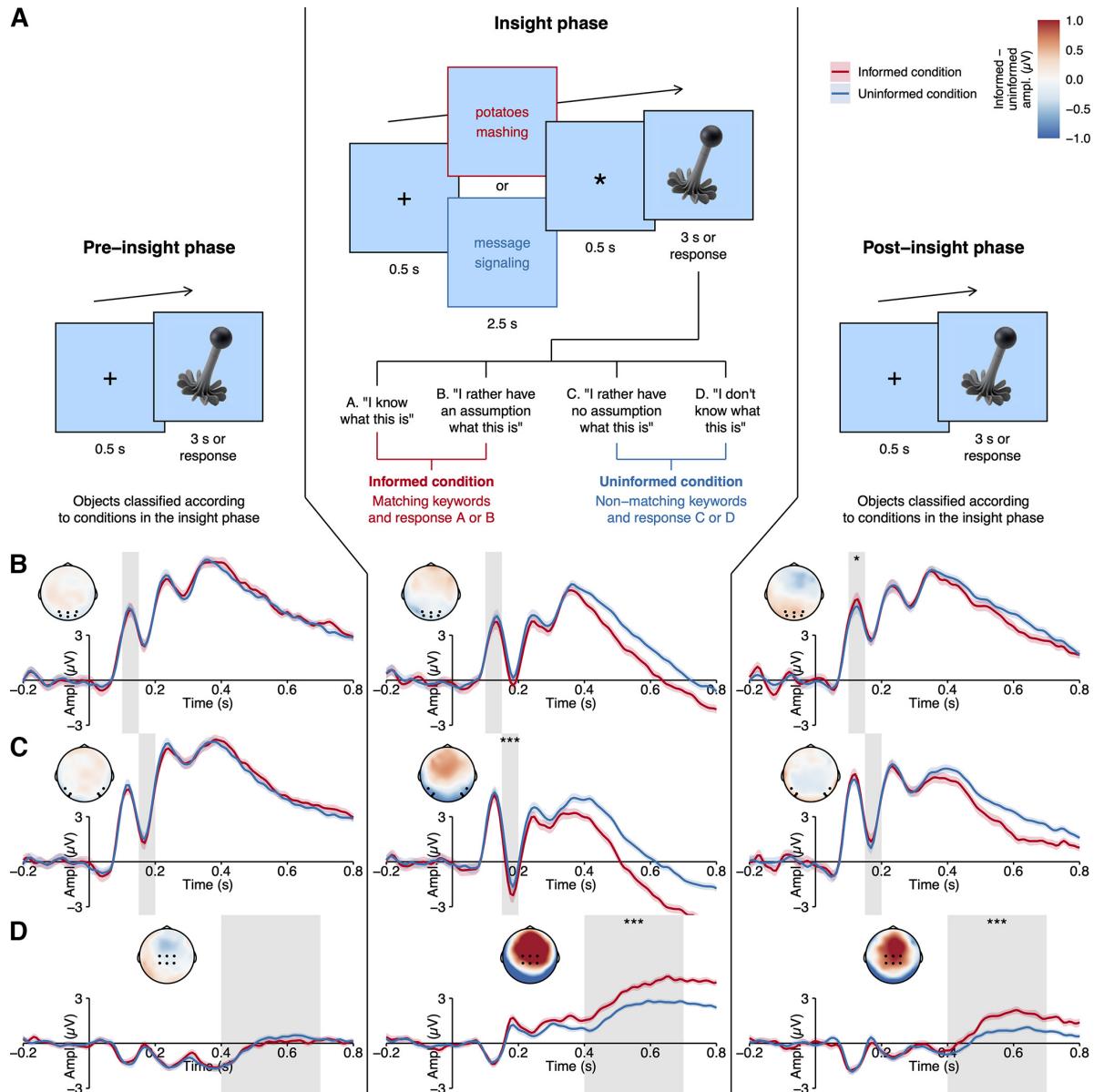


Figure 1. Experimental design and ERP results contrasting semantically informed perception and uninformed perception. **A**, In the pre-insight phase, participants were presented with 120 unfamiliar objects and indicated whether they knew what kind of object they were viewing. In the insight phase, half of these objects were presented with matching keywords (in red for illustration), leading to semantically informed perception, and the other half with nonmatching keywords (in blue for illustration), leading to uninformed perception. In the post-insight phase, the same objects were presented again without keywords. **B-D**, ERP waveforms and scalp topographies for the P1 component (**B**), for the N170 component (**C**), and for the N400 component (**D**) for objects with semantically informed perception versus uninformed perception within the three different phases. Semantically informed perception was associated with more negative amplitudes in the N170 component during the insight phase, less negative amplitudes in the N400 component during the insight and post-insight phases, and more positive amplitudes in the P1 component during the post-insight phase. Waveform plots represent the ERP amplitudes averaged across channels in the ROIs (P1: P03, P04, POz, O1, O2, Oz; N170: P7, P8, PO7, PO8, PO9, PO10; N400: C1, C2, Cz, CP1, CP2, CPz; see black dots in the scalp topographies). Colored ribbons around the ERP waveforms represent ± 1 SEM across participants. Topographies represent the difference in ERP amplitudes at all channels on the scalp, averaged across the time windows of interest (P1: 100–150 ms; N170: 150–200 ms; N400: 400–700 ms; see gray areas in the ERP waveforms). Ampl., Amplitude. * $p < 0.05$. *** $p < 0.001$.

Participants were asked to respond as quickly and as accurately as possible by pressing one of four buttons with the index or middle finger of their left or right hand, respectively. The mapping of the rating scale to the four buttons (left to right or right to left) was counterbalanced across participants.

In the insight phase, the 120 unfamiliar objects were presented for a second time, now preceded either by matching keywords (leading to

semantically informed perception) or by nonmatching keywords (leading to uninformed perception). Each trial consisted of a fixation cross presented for 0.5 s, followed by the presentation of the keywords for 2.5 s. Then, an asterisk was presented in the middle of the screen for another 0.5 s, followed by the presentation of the object until a response was made or until a timeout after 3 s. The objects were presented in blocks of 30 trials so that within each block there were 15 objects from

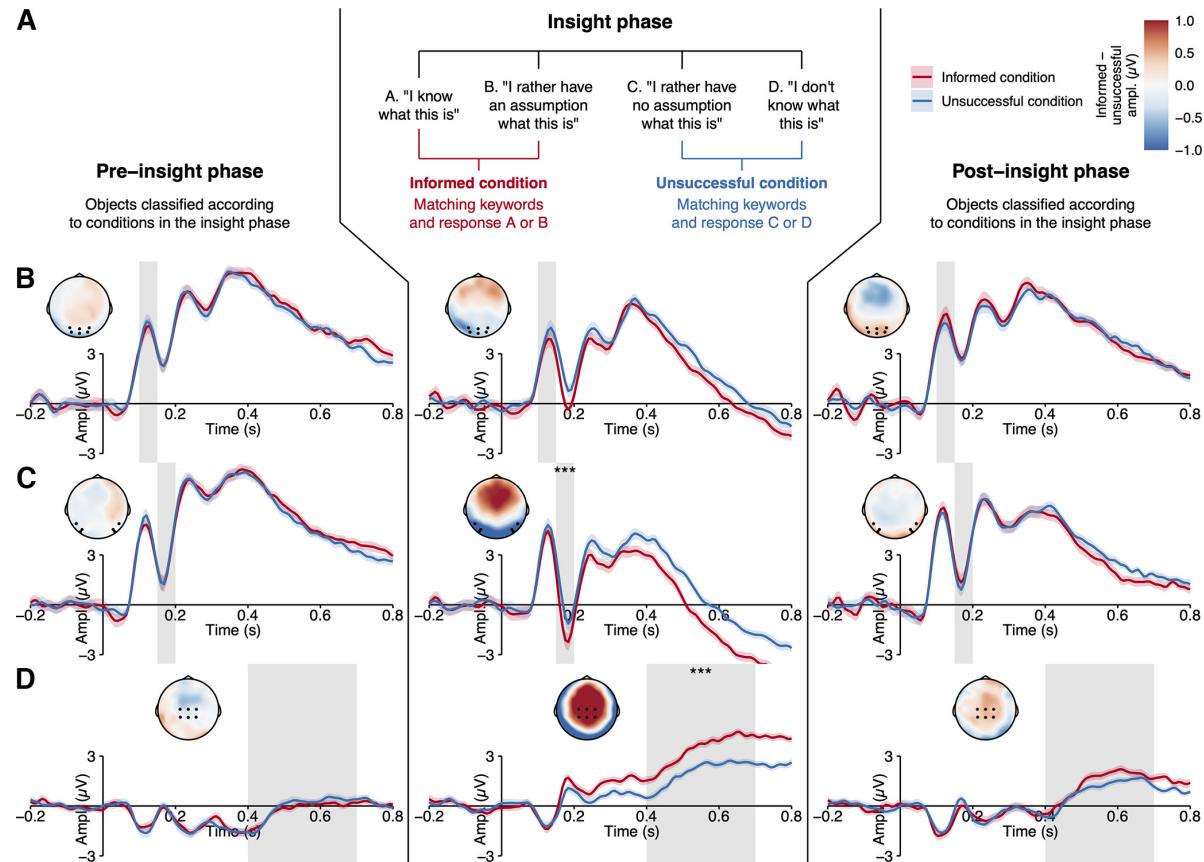


Figure 2. ERP results contrasting semantically informed perception and unsuccessfully informed perception. **A**, Assignment of objects to conditions based on participants' response to unfamiliar objects presented with matching keywords. **B–D**, ERP waveforms and scalp topographies for the P1 component (**B**), for the N170 component (**C**), and for the N400 component (**D**) for objects with semantically informed perception versus unsuccessfully informed perception within the three different phases. Semantically informed perception was associated with more negative amplitudes in the N170 component and less negative amplitudes in the N400 component during the insight phase. Waveform plots represent the ERP amplitudes averaged across channels in the ROIs (P1: P03, P04, POz, O1, O2, Oz; N170: P7, P8, PO7, PO8, PO9, PO10; N400: C1, C2, Cz, CP1, CP2, CPz; see black dots in the scalp topographies). Colored ribbons around the ERP waveforms represent ± 1 SEM across participants. Topographies represent the difference in ERP amplitudes at all channels on the scalp, averaged across the time windows of interest (P1: 100–150 ms; N170: 150–200 ms; N400: 400–700 ms; see gray areas in the ERP waveforms). Ampl., Amplitude. *** $p < 0.001$.

each of the two experimental conditions and so that objects were heterogeneous in terms of their shape, visual complexity, and functional category (e.g., medical devices, musical instruments).

In the post-insight phase, the unfamiliar objects were presented for a third time, with the same trial structure as in the pre-insight phase, that is, without any keywords. The insight and post-insight phases were presented in an interleaved fashion so that, after the presentation of one block of 30 objects in the insight phase (with keywords), participants took a self-timed break and continued with the same block of 30 objects in the post-insight phase (without keywords) before moving on to the next block consisting of 30 different objects. They continued like this until all four blocks were completed in both phases. In total, the experiment consisted of 480 trials (120 familiar objects in the pre-insight phase and 120 unfamiliar objects in the pre-insight, insight, and post-insight phases). Participants took approximately 35 min to complete the experiment.

Behavioral data analysis. We used participants' behavioral responses to verify our experimental manipulation and to assign each object to a semantic condition. First, we used participants' responses from the pre-insight phase (when objects were presented for the first time, without keywords) to make sure that objects were indeed unfamiliar to them. That is, we excluded any objects for which participants responded with "I know what this is" in the pre-insight phase. Next, we used participants' responses from the insight phase (when objects were presented for the second time, preceded by keywords) to assign the remaining

objects to different semantic conditions: Semantically informed perception, uninformed perception, and unsuccessfully informed perception. The semantically informed condition consisted of objects that were presented with matching keywords and for which the participant responded with knowing what the object was or having an assumption. The uninformed condition consisted of objects that were presented with nonmatching keywords and for which the participant responded with not knowing what the object was or having rather no assumption. The unsuccessfully informed condition consisted of objects that were presented with matching keywords but for which the participant responded with not knowing what the object was or having rather no assumption. This assignment of objects to semantic conditions (informed, uninformed, or unsuccessfully informed) was carried over from the insight phase to the other two phases (pre-insight and post-insight). This allowed us to test, on the one hand, if the objects differed in important aspects even before any keywords were presented (pre-insight phase) and, on the other hand, if the semantic information acquired in the insight phase had any downstream effects on the subsequent perception of the objects (post-insight phase).

EEG recording and preprocessing. The continuous EEG was recorded from 62 Ag/AgCl scalp electrodes placed according to the extended 10–20 system (American Electroencephalographic Society, 1991) and referenced online to an external electrode placed on the left mastoid (M1). Two additional external electrodes were placed on the right mastoid

(M2) and below the left eye (IO1), respectively. Electrode impedances were kept below 5 k Ω . An online bandpass filter with a high-pass time-constant of 10 s (0.016 Hz) and a low-pass cutoff frequency of 1000 Hz was applied before digitizing the signal at a sampling rate of 500 Hz.

The data were preprocessed offline using custom functions (available at <https://github.com/alexenge/hu-neuro-pipeline/tree/v0.6.4>) based on the MNE-Python software (version 1.3.0) (Gramfort et al., 2013) in Python (version 3.8) (Van Rossum and Drake, 2009). First, the data were downsampled to 125 Hz and rereferenced to the common average of all scalp channels. Next, artifacts resulting from blinks and eye movements were removed via independent component analysis (ICA) on a high-pass filtered copy of the data (cutoff = 1 Hz). A variable number of components per participant were extracted from an initial PCA so that they explained at least 99% of the variance in the data (mean = 28.85 components, range 16–40). Then the ICA was fitted based on these components using the FastICA algorithm (Hyvärinen, 1999). Any independent components showing significant correlations with either of two virtual EOG channels (VEOG: IO1-Fp1; HEOG: F9-F10) were removed automatically using MNE-Python's *find_bads_eog* method. This was the case for an average of 2.38 components per participant (range = 2–4 components).

For the analysis of ERP amplitudes, a zero-phase, noncausal FIR filter with a lower pass-band edge at 0.1 Hz (transition bandwidth: 0.1 Hz) and an upper pass-band edge at 30 Hz (transition bandwidth: 7.5 Hz) was applied. Next, the continuous EEG was segmented into epochs ranging from −500 to 1500 ms relative to the onset of the presentation of each unfamiliar object. These epochs were baseline-corrected by subtracting the average voltage during the interval of −200 to 0 ms relative to stimulus onset. Epochs containing artifacts despite ICA, defined as peak-to-peak amplitudes exceeding 150 μ V, were removed from further analysis. This led to the exclusion of an average of 11 trials per participant (= 3.1%; range 0–109 trials). Single-trial ERPs were computed as the mean amplitude across time windows and ROIs defined *a priori*, namely, 100–150 ms at channels PO3, PO4, POz, O1, O2, and Oz and for the P1 component, 150–200 ms at channels P7, P8, PO7, PO8, PO9, and PO10 for the N170 component, and 400–700 ms at channels C1, C2, Cz, CP1, CP2, and CPz for the N400 component. We chose a later time window for the N400 component than what is typically used in experiments with verbal materials (e.g., 300–500 ms), in accordance with a previously published dataset, which showed that the 400–700 ms time window is most robustly associated with the semantic processing of visual objects (Kovalenko et al., 2012).

Statistical analysis. The resulting mean ERP amplitudes were analyzed on the single-trial level using linear mixed-effects regression models because these models allow to control for repeated measures of participants and stimuli, while also being robust against an unbalanced number of trials per condition (Bürki et al., 2018; Frömer et al., 2018; Brown, 2021). We computed three models predicting P1, N170, and N400 amplitudes, respectively. All models included the following fixed effects of interest: (1) the phase of the experiment, coded as a repeated contrast (i.e., subtracting the first phase from the second phase and the second phase from the third phase, the intercept being the grand mean across all three phases) (Schad et al., 2020); (2) the condition of the object, coded as a custom contrast (i.e., subtracting the uninformed condition from the informed condition and the unsuccessfully informed condition from the informed condition, the intercept being the grand mean across all three conditions); and (3) the two-way interaction of phase and condition. We also added the results from the online pre-rating study (mean cosine distance between rating study-generated keywords and keywords presented in the main experiment) as an additional covariate of no interest (see Online Fig. 2). This was to control for the possibility that participants might have been partly familiar with some of the objects and/or able to guess their function from their visual appearance alone. For the random effects, we determined the most parsimonious structure supported by the data using the automatic procedure proposed by Matuschek et al. (2017). This involved starting with a maximal model that contained all random parameters (intercepts, slopes, and correlations) and then iteratively removing terms as long as this did not

result in a significant drop in model fit (likelihood ratio test, p value cutoff = 0.20; for the final model syntax and model outputs, see Online Results 1). All models were fitted in R (version 4.2.1) (R Core Team, 2022) using the lme4 package (version 1.1.30) (Bates et al., 2015) with the optimizer function *bobyqa* and a maximum of 10^6 iterations for maximum likelihood estimation. The model selection algorithm via likelihood ratio tests was performed using the buildmer package (version 2.7) (Voeten, 2022).

To investigate whether semantically informed perception had an influence on the ERPs within each phase of the experiment, we calculated pairwise comparisons contrasting the semantically informed condition against the uninformed condition within the pre-insight, insight, and post-insight phases. In the same way, we computed pairwise comparisons contrasting the semantically informed condition against the unsuccessfully informed condition. This was done using the emmeans package (version 1.8.2) (Lenth, 2022) and with Bonferroni correction for the three phases of the experiment. All p values were computed by approximating the relevant denominator degrees of freedom using Satterthwaite's method as implemented in the lmerTest package (version 3.1.3) (Kuznetsova et al., 2017).

Time-frequency analysis. For our exploratory analysis of event-related power, we first created new epochs from the ICA-corrected but unfiltered raw data. Epochs that were marked as bad for the ERP analysis (i.e., with peak-to-peak amplitudes exceeding 150 μ V) were also removed from the time-frequency analysis. The remaining epochs were convolved with a family of Morlet wavelets, increasing linearly in their frequency from 4 to 40 Hz in steps of 1 Hz and in their width from 2 cycles to 20 cycles in steps of 0.5 cycles. To adjust for the typical 1/ f shape of the EEG, the power values were transformed into percent signal change by first subtracting and then dividing by the average power at each frequency over the entire epoch (Grandchamp and Delorme, 2011). We then performed baseline correction by subtracting the average power during the pre-stimulus interval from −450 to −50 ms relative to object onset.

For statistical analysis of event-related power, we conducted cluster-based permutation tests (Maris and Oostenveld, 2007), separately for each of the three phases of the experiment (pre-insight, insight, and post-insight). First, we averaged trials belonging to the same condition and then subtracted these average responses from one another to compute the difference between the semantically informed condition and each of two control conditions (uninformed or unsuccessfully informed) for each participant. We then conducted one-sample *t* tests in a mass-univariate fashion and grouped significant results (cluster-forming threshold, $p < 0.05$) into clusters if they occurred at neighboring time points, frequencies, or channels. Neighboring channels were defined using the Delaunay triangulation based on 2D electrode locations as implemented in MNE-Python. To obtain family-wise error-corrected p values at the cluster level, we compared the cluster mass (i.e., the sum of the *t* values) of each observed cluster to an empirical null distribution of cluster masses obtained from 5000 permutations with random sign flips. Clusters were considered to be statistically significant if their mass exceeded the 98.33rd percentile of this distribution (i.e., cluster-level threshold, $p < 0.05$, Bonferroni-corrected for the three phases of the experiment).

Data and code accessibility. The EEG data are available on reasonable request from the corresponding author because we had not asked participants for their consent to make the data publicly available. The experimental stimuli (object images and keywords) and the code for data analysis are available at <https://doi.org/10.17605/osf.io/ukscb>.

In addition to the software mentioned above, our code relies on the tidyverse set of R packages (version 1.3.2) (Wickham et al., 2019) for data wrangling, the ggplot2 (version 3.3.6) (Wickham, 2016), cowplot (version 1.1.1) (Wilke, 2020), and eegUtils (version 0.7.0) (Craddock, 2022) packages for visualization; the papaja package (version 0.1.1) (Aust and Barth, 2022) for statistical reporting; and the LSAfun package (version 0.6.3) (Günther et al., 2015) for the latent semantic analysis of the online rating study data. We used the workflow developed by Peikert and Brandmaier (2021) to ensure the long-term reproducibility of our analysis pipeline.

Results

Behavioral data

Based on the experimental manipulation (matching or non-matching keywords) and each individual participant's behavioral response (positive or negative regarding knowing what the object was), we assigned 26.67 ± 10.51 (mean \pm SD) objects to the semantically informed condition (i.e., matching keywords and positive response), 48.98 ± 7.71 objects to the uninformed condition (i.e., nonmatching keywords and negative response), and 26.77 ± 10.48 objects to the unsuccessfully informed condition (i.e., matching keywords but negative response). This assignment of objects to conditions was based solely on the insight phase of the experiment, when objects were presented with keywords for the first time, and carried over to analyze the data from the pre-insight and post-insight phases as well. The remaining objects were excluded because of an implausible response pattern (i.e., nonmatching keywords but positive response; 5.29 ± 5.78 objects), because of being known to the participant in the pre-insight phase (i.e., before any keywords had been presented; 7.10 ± 5.00 objects), or because of technical errors (i.e., reaction times of 0 ms recorded by the presentation software, 5.19 ± 5.84 objects).

Event-related potentials

Averaged across conditions, P1, N170, and N400 amplitudes differed as a function of the phase of the experiment (pre-insight, insight, or post-insight, all F values > 14.28 , all p values < 0.001). In addition, N400 amplitudes differed as a function of the condition (semantically informed, uninformed, or unsuccessfully informed), averaged across the three phases of the experiment ($F_{(2, 14,180)} = 22.91$, $p < 0.001$). Crucially, the phase \times condition interaction was significant for all three ERP components (all F values > 2.54 , all p values < 0.038). To answer our main research question, we decomposed these interactions into pairwise comparisons between semantic conditions within each of the three phases of the experiment.

In the pre-insight phase, when objects were unfamiliar to participants and presented without keywords, no differences emerged between the semantically informed condition and the uninformed condition in any of the ERP components (all $|t|$ values < 0.86 , all p values > 0.999 ; see Fig. 1*B–D*). Likewise, there were no differences between the semantically informed condition and the unsuccessfully informed condition in any of the ERP components (all $|t|$ values < 0.55 , all p values > 0.999 ; see Fig. 2*B–D*). This was expected given that the keywords that would provide additional semantic information had not yet been presented and given that the assignment of objects to conditions was counterbalanced across participants as to control for low-level visual differences.

In the insight phase, half of the unfamiliar objects were presented with matching keywords, leading to semantically informed perception, and the other half were presented with nonmatching keywords, keeping the perception of the objects semantically uninformed. Semantically informed perception in this phase was associated with enlarged (i.e., more negative) amplitudes in the N170 component ($b = -0.64 \mu\text{V}$, $t_{(14,194)} = -3.63$, $p = 0.001$) and reduced (i.e., less negative) amplitudes in the N400 component ($b = 1.09 \mu\text{V}$, $t_{(14,156)} = 7.17$, $p < 0.001$). There was no reliable difference between conditions in the P1 component ($b = -0.27 \mu\text{V}$, $t_{(14,138)} = -1.49$, $p = 0.406$). The same effects were found when comparing objects with semantically informed perception to those objects that were also presented with matching keywords but for which participants

still indicated not knowing what the object was (i.e., the unsuccessfully informed condition). Compared with these objects, semantically informed perception was associated with enlarged (i.e., more negative) amplitudes in the N170 component ($b = -0.90 \mu\text{V}$, $t_{(14,203)} = -4.37$, $p < 0.001$) and reduced (i.e., less negative) amplitudes in the N400 component ($b = 1.22 \mu\text{V}$, $t_{(13,925)} = 7.00$, $p < 0.001$), while there was no reliable difference in the P1 component ($b = -0.39 \mu\text{V}$, $t_{(13,860)} = -1.84$, $p = 0.196$).

In the post-insight phase, the unfamiliar objects were presented for a third time and without any keywords, mirroring the pre-insight phase. As in the insight phase, semantically informed perception was associated with reduced (i.e., less negative) amplitudes in the N400 component ($b = 0.80 \mu\text{V}$, $t_{(14,031)} = 5.23$, $p < 0.001$), while the effect in the N170 component did not recur ($b = 0.26 \mu\text{V}$, $t_{(13,742)} = 1.46$, $p = 0.431$). Instead, the P1 component was significantly enlarged (i.e., more positive) in response to objects for which semantically informed perception had taken place ($b = 0.46 \mu\text{V}$, $t_{(13,047)} = 2.52$, $p = 0.036$). These effects did not occur when comparing the semantically informed condition to the unsuccessfully informed condition, with no reliable differences in the P1, N170, or N400 components (all $|t|$ values < 2.17 , all p values > 0.090).

Event-related power

In an exploratory time-frequency analysis, we checked for differences in event-related power between semantically informed perception and uninformed perception within each of the three phases of the experiment. Cluster-based permutation tests (Maris and Oostenveld, 2007) revealed no significant clusters in the pre-insight phase (see Online Fig. 3, all p values > 0.789), but one significant cluster in the insight phase (Fig. 3, $p_{\text{cluster}} = 0.002$) and one significant cluster in the post-insight phase (Fig. 4, $p_{\text{cluster}} = 0.002$). These two significant clusters in the insight and post-insight phases were similar in their direction, latency, frequency range, and topographic distribution. Both clusters had a negative sign, started at approximately 600 ms after object onset (but see Sassenhagen and Draschkow, 2019) and continued all the way until the end of the analyzed period at 1400 ms. They spanned a broad range of frequencies in the alpha and lower beta range as well as a broad set of channels but appeared to be most focal at approximately 15 Hz and parietal channels. Thus, semantically informed perception seems to alter not only early, evoked activity (see Event-related potentials) but also later, induced activity, in the form of a reduction of post-stimulus power at parietal channels in the range of alpha and lower beta frequencies.

We repeated this analysis to look for differences between semantically informed perception and unsuccessfully informed perception within each of the three phases of the experiment. There were no significant clusters in the pre- or post-insight phases (see Online Figs. 4 and 6; all p values > 0.216) but one significant cluster in the insight phase (see Online Fig. 5; $p_{\text{cluster}} = 0.016$). This cluster was similar to the ones described above in terms of its latency, frequency range, and topographic distribution.

Discussion

We found that providing participants once with semantic information about previously unfamiliar objects instantly led to enlarged (i.e., more negative) ERP amplitudes in the N170 component and reduced (i.e., less negative) ERP amplitudes in the N400 component. When the same objects were presented again, the N400 component remained reduced and the P1 component

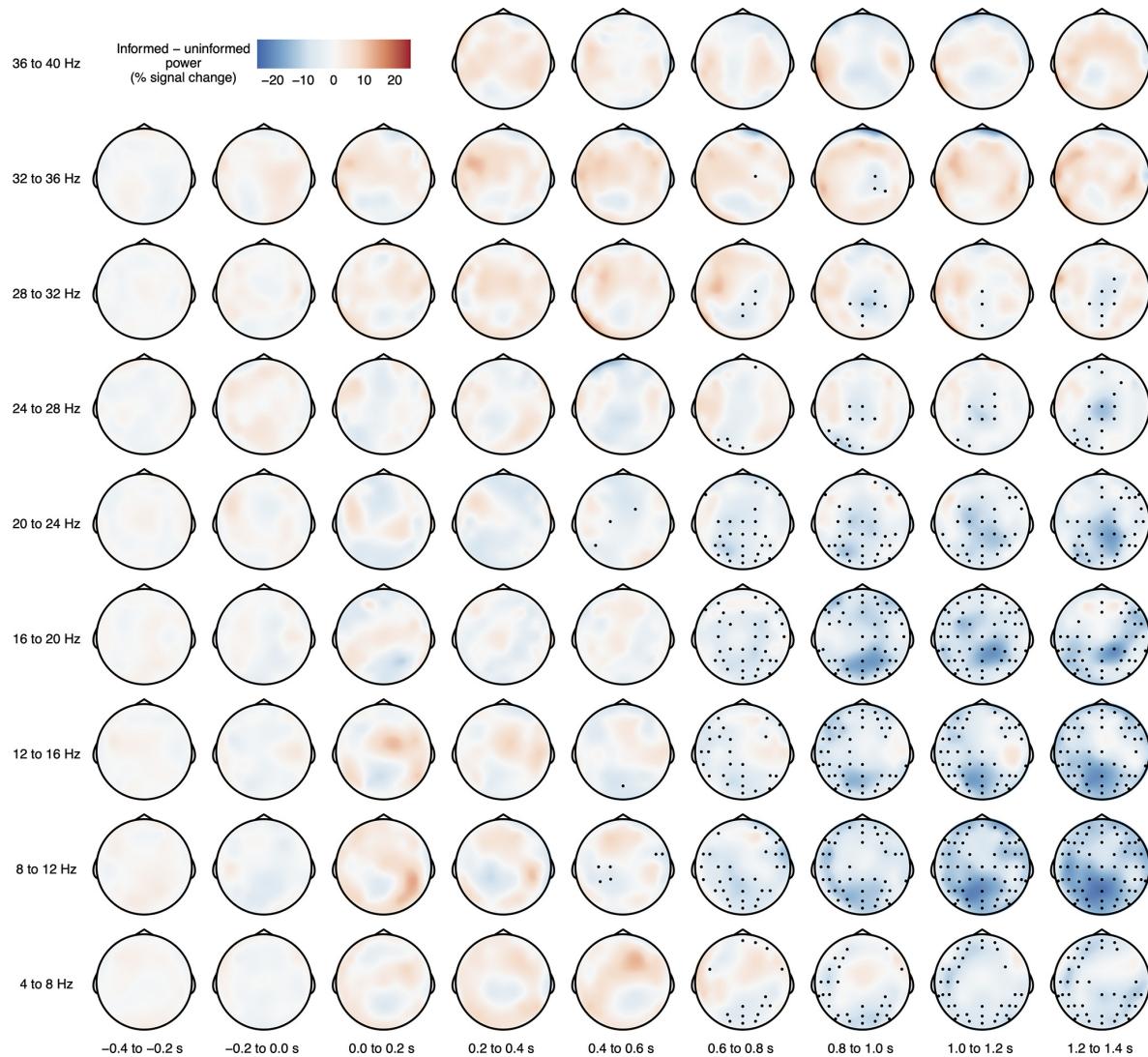


Figure 3. Time-frequency results for the insight phase. Each topographic plot represents the difference in event-related power (in units of percent signal change) between the semantically informed condition and the uninformed condition, grand-averaged across participants. Black dots represent EEG channels that were part of a cluster for which this difference was statistically significant ($p_{\text{cluster}} = 0.002$).

was now enlarged (i.e., more positive) in response to objects that had previously triggered semantically informed perception. An exploratory time-frequency analysis revealed that semantically informed perception was accompanied by a late reduction in event-related power in the alpha and lower beta ranges.

The N400 effect indicates that acquiring an understanding of the object lessened participants' demand for semantic processing (Kutas and Federmeier, 2011). It replicates previous work showing larger N400 amplitudes for pictures when they are difficult to understand in and of themselves (e.g., Supp et al., 2005; Abdel Rahman and Sommer, 2008) or difficult to integrate into the preceding context (e.g., Barrett and Rugg, 1990; Ganis et al., 1996; Hirschfeld et al., 2011). The latency of this effect suggests a post-perceptual locus in the semantic system.

Our exploratory time-frequency analysis revealed a late (>600 ms) reduction in power in the alpha and lower beta ranges (approximately 8–20 Hz). Like the N400 effect, this occurred

as soon as participants had received the semantic information and recurred once the objects were re-encountered without any semantic information. Reductions in alpha/beta power have been shown to correlate with clearer representations of stimulus-specific information, as measured using representational similarity analysis (Griffiths et al., 2019), and with successfully forming new semantic memories (Hanslmayr et al., 2009). This may be because of a dampening of alpha/beta oscillations, which creates favorable conditions for high-level cortical information processing and encoding.

In contrast to the N400 and event-related power, the N170 was modulated only on those initial trials on which the relevant semantic information was presented directly before the object. It therefore constitutes an online marker of semantic insight, that is, of participants suddenly understanding the visual objects in the light of the information provided by the keywords. The N170 is associated with the holistic perception of faces (Sagiv and Bentin, 2001; Eimer et al., 2011) and other stimuli of visual

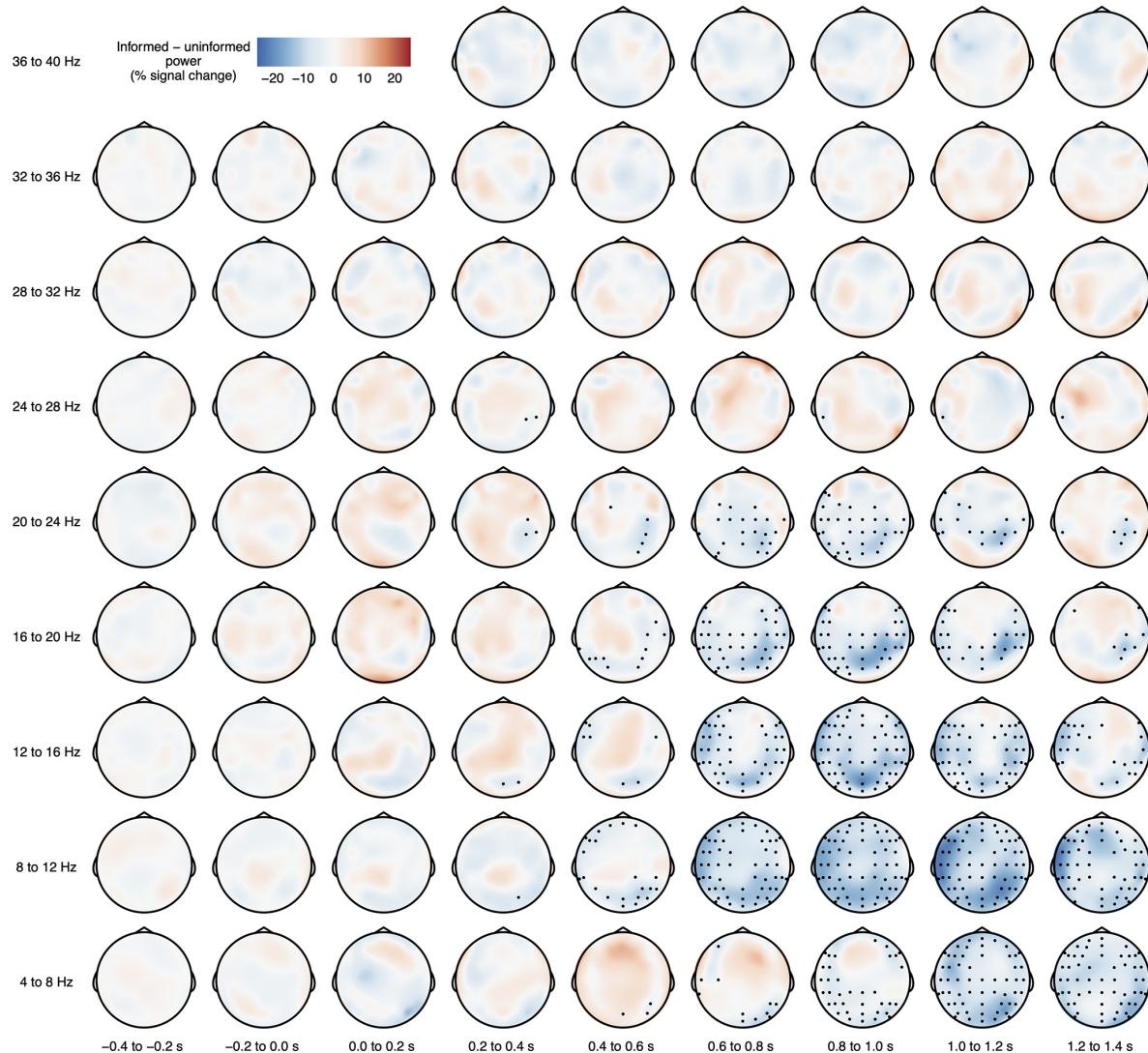


Figure 4. Time-frequency results for the post-insight phase. Each topographic plot represents the difference in event-related power (in units of percent signal change) between the semantically informed condition and the uninformed condition, grand-averaged across participants. Black dots represent EEG channels that were part of a cluster for which this difference was statistically significant ($p_{\text{cluster}} = 0.002$).

expertise (Tanaka and Curran, 2001; Rossion et al., 2002); that is, it is sensitive to factors that go beyond structural encoding and categorical perception (Thierry et al., 2007a,b; Dering et al., 2011). It being enlarged may reflect that the semantic information made participants experience the configuration of the visual features of the objects in a new and meaningful way. This is supported by previous findings of enlarged N170 amplitudes for scrambled face stimuli after participants had been shown the original version of the face (Bentin and Golland, 2002), as well as for line drawings of meaningful objects compared with nonobjects (Beaucousin et al., 2011). Together, this suggests an online impact of meaningfulness on the higher-level perception of visual objects, integrating across their visual features.

The P1, unlike the N400 and N170, was modulated by semantic information only one trial after this information had been obtained. This replicates previous studies showing modulations of the P1 when participants had learned meaningful information

about previously unfamiliar objects (Abdel Rahman and Sommer, 2008; Maier and Abdel Rahman, 2018, 2019; Samaha et al., 2018; Weller et al., 2019). The present study adds that the P1 effect does not take an extensive learning history to develop; instead, it can be observed as soon as one trial after semantic insight has happened. Because the P1 is associated with lower-level sensory processing (Johannes et al., 1995; Pratt, 2011; Luck, 2014), we take its susceptibility to semantic information as an indicator that knowledge about the function of an object can change how we perceive its low-level features (Athanasopoulos and Casaponsa, 2020).

The P1 and N170 were modulated in different phases of our study, suggesting that they reflect different aspects of top-down processing with different time courses and neuroanatomical implementations. The time course of the N170 is consistent with a top-down influence of nonvisual areas in the prefrontal and temporo-parietal cortices on visual areas, whereas modulations of the P1 component may reflect recurrent processing within the

visual system (Wyatte et al., 2014). Here we could show that the former pathway seems to be able to convey semantic information instantaneously (i.e., within the same trial), whereas the latter pathway seems to take at least one additional encounter with the object to emerge. While the limited spatial resolution of the EEG precludes localization, there is converging fMRI and psychophysical evidence that semantic information can feed back into areas in the lateral occipital cortex as well as early retinotopic cortex (areas V1, V2, and V3) (Hsieh et al., 2010; Clarke et al., 2016; Teufel et al., 2018), consistent with the neural generators of the N170 and P1 in the ERP.

The top-down modulation of visual ERPs by semantic information challenges a modular view of visual perception (Fodor, 1983; Pylyshyn, 1999). Proponents of this view have pointed out important shortcomings of previous studies that had claimed to demonstrate top-down effects of cognition on perception (Machery, 2015; Firestone and Scholl, 2016). We addressed as many of these shortcomings as possible: We ensured that there were no visual differences between conditions (with counterbalancing and the pre-insight phase as a negative control), we used ERPs as an objective and time-resolved measure to disentangle perceptual and postperceptual effects, and we reduced response and demand biases by keeping the manipulation (i.e., matching or nonmatching keywords) obscure to participants and by including well-known objects as filler stimuli.

One could argue that the effects presented here might be reducible to more basic mechanisms, such as semantic priming. Indeed, the keywords that were presented before each object in the semantically informed condition were chosen such that they matched the function of the object and often had a direct relationship to certain visual features of the object. This might have induced semantic priming, which is supported by the reduction in N400 amplitudes (e.g., Bentin et al., 1985; Kellenbach et al., 2000). However, there are at least three arguments why semantic priming cannot account for our main findings, that is, the influence of semantic information on the P1 and N170. First, for both components, ERP amplitudes were enlarged (i.e., more positive for the P1 and more negative for the N170) during semantically informed perception, whereas semantic priming typically leads to reduced ERP amplitudes. Second, these effects were not just observed when comparing semantically informed perception (with matching keywords) and uninformed perception (with nonmatching keywords), but also when comparing semantically informed perception with unsuccessfully informed perception. In the latter case, all objects were preceded by keywords that matched the function of the object to a similar degree, making semantic priming less likely. Third, the results from an online rating study (see Materials and Methods; Online Fig. 1) indicated that people by and large did not spontaneously associate the unfamiliar objects with a particular function, and also allowed us to statistically control for the closeness of peoples' guesses and the true function of each object.

A theoretical framework that would explicitly predict the observed P1 and N170 effects in our study is lacking at present. However, the effects are consistent with the reverse hierarchy theory (Ahissar and Hochstein, 2004), which posits that objects first enter visual consciousness at an abstract, conceptual level. Once this initial "vision at a glance" has taken place, feedback connections to earlier layers of the visual system are being accessed to extract the relevant lower-level features ("vision with scrutiny"). This reverse trajectory down the visual hierarchy may explain (1) the semantically induced changes to the fMRI signal

in lateral occipital cortex and retinotopic cortex (e.g., Hsieh et al., 2010) as well as (2) the modulations of early visual ERP components observed in the present study and others (e.g., Abdel Rahman and Sommer, 2008; Maier et al., 2014; Samaha et al., 2018). An important role of top-down mechanisms for object recognition is also posited by theories of predictive coding and Bayesian inference (e.g., Yuille and Kersten, 2006; Xu and Tenenbaum, 2007; Clark, 2013; Panichello et al., 2013; Lupyan, 2015). Despite the theoretical advances, detailed descriptions of these top-down effects at the algorithmic and implementational levels remain a challenge for future work.

In conclusion, the present study provides preliminary evidence that, whenever we receive semantic information about a previously unfamiliar object, this information has an immediate influence on our visual processing of this object. The immediacy of this influence is remarkable in at least two different ways: First, it does not require an extensive learning history but can instead be observed within the same trial in which the information has been presented and/or a single trial later. Second, the time course of this influence suggests that it manifests itself not only at later, postperceptual stages (>400 ms), typically associated with semantic processing, but also at much earlier stages within the first 200 ms, associated with visual perception itself.

References

- Abdel Rahman R, Sommer W (2008) Seeing what we know and understand: how knowledge shapes perception. *Psychon Bull Rev* 15:1055–1063.
- Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci* 8:457–464.
- American Electroencephalographic Society (1991) American Electroencephalographic Society guidelines for standard electrode position nomenclature. *J Clin Neurophysiol* 8:200–202.
- Athanopoulos P, Casaponsa A (2020) The Whorfian brain: neuroscientific approaches to linguistic relativity. *Cogn Neuropsychol* 37:393–412.
- Aust F, Barth M (2022) papaja: prepare reproducible APA journal articles with R markdown. <https://github.com/crsh/papaja>.
- Barrett SE, Rugg MD (1990) Event-related potentials and the semantic matching of pictures. *Brain Cogn* 14:201–212.
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Soft* 67:1–48.
- Beaucousin V, Cassotti M, Simon G, Pineau A, Kostova M, Houdé O, Poirel N (2011) ERP evidence of a meaningfulness impact on visual global/local processing: when meaning captures attention. *Neuropsychologia* 49:1258–1266.
- Bentin S, Golland Y (2002) Meaningful processing of meaningless stimuli: the influence of perceptual experience on early visual processing of faces. *Cognition* 86:B1–B14.
- Bentin S, McCarthy G, Wood CC (1985) Event-related potentials, lexical decision and semantic priming. *Electroencephalogr Clin Neurophysiol* 60:343–355.
- Brown VA (2021) An introduction to linear mixed-effects modeling in R. *Adv Methods Pract Psychol Sci* 4:1–19.
- Bürki A, Frossard J, Renaud O (2018) Accounting for stimulus and participant effects in event-related potential analyses to increase the replicability of studies. *J Neurosci Methods* 309:218–227.
- Churchland PS, Ramachandran VS, Sejnowski TJ (1994) A critique of pure vision. In: Computational neuroscience: large-scale neuronal theories of the brain. (Koch C, Davis JL, eds), pp 23–60. Cambridge, MA: Massachusetts Institute of Technology.
- Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 36:181–204.
- Clarke A, Pell PJ, Ranganath C, Tyler LK (2016) Learning warps object representations in the ventral temporal cortex. *J Cogn Neurosci* 28:1010–1023.
- Craddock M (2022) eegUtils: utilities for electroencephalographic (EEG) analysis. <https://github.com/craddm/eegUtils>.
- Dering B, Martin C, Moro S, Pegna A, Thierry G (2011) Face-sensitive processes one hundred milliseconds after picture onset. *Front Hum Neurosci* 5:93.

- DiCarlo JJ, Zoccolan D, Rust NC (2012) How does the brain solve visual object recognition? *Neuron* 73:415–434.
- Eimer M, Gosling A, Nicholas S, Kiss M (2011) The N170 component and its links to configural face processing: a rapid neural adaptation study. *Brain Res* 1376:76–87.
- Firestone C, Scholl BJ (2016) Cognition does not affect perception: evaluating the evidence for ‘top-down’ effects. *Behav Brain Sci* 39:e229.
- Fodor JA (1983) The modularity of mind. Cambridge, MA: Massachusetts Institute of Technology.
- Friston K, Kiebel S (2009) Predictive coding under the free-energy principle. *Philos Trans R Soc Lond B Biol Sci* 364:1211–1221.
- Frömer R, Maier M, Abdel Rahman R (2018) Group-level EEG-processing pipeline for flexible single trial-based analyses including linear mixed models. *Front Neurosci* 12:48.
- Ganis G, Kutas M, Sereno MI (1996) The search for ‘common sense’: an electrophysiological study of the comprehension of words and pictures in reading. *J Cogn Neurosci* 8:89–106.
- Gauthier I, James TW, Curby KM, Tarr MJ (2003) The influence of conceptual knowledge on visual discrimination. *Cogn Neuropsychol* 20:507–523.
- Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hämäläinen M (2013) MEG and EEG data analysis with MNE-Python. *Front Neurosci* 7:267.
- Grandchamp R, Delorme A (2011) Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. *Front Psychol* 2:236.
- Griffiths BJ, Mayhew SD, Mullinger KJ, Jorge J, Charest I, Wimber M, Hanslmayr S (2019) Alpha/beta power decreases track the fidelity of stimulus-specific information. *Elife* 8:e49562.
- Günther F, Dudschig C, Kaup B (2015) LSAfun: an R package for computations based on Latent Semantic Analysis. *Behav Res Methods* 47:930–944.
- Günther F, Rinaldi L, Marelli M (2019) Vector-space models of semantic representation from a cognitive perspective: a discussion of common misconceptions. *Perspect Psychol Sci* 14:1006–1033.
- Hanslmayr S, Spitzer B, Bäuml KH (2009) Brain oscillations dissociate between semantic and nonsemantic encoding of episodic memories. *Cereb Cortex* 19:1631–1640.
- Hirschfeld G, Zwitslerlood P, Dobel C (2011) Effects of language comprehension on visual processing MEG dissociates early perceptual and late N400 effects. *Brain Lang* 116:91–96.
- Hsieh PJ, Vul E, Kanwisher N (2010) Recognition alters the spatial pattern of fMRI activation in early retinotopic cortex. *J Neurophysiol* 103:1501–1507.
- Hyvärinen A (1999) Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw* 10:626–634.
- Johannes S, Münte TF, Heinze HJ, Mangun GR (1995) Luminance and spatial attention effects on early visual processing. *Brain Res Cogn Brain Res* 2:189–205.
- Kellenbach ML, Wijers AA, Mulder G (2000) Visual semantic features are activated during the processing of concrete words: event-related potential evidence for perceptual semantic priming. *Brain Res Cogn Brain Res* 10:67–75.
- Kovalenko LY, Chaumon M, Busch NA (2012) A pool of pairs of related objects (POPORO) for investigating visual semantic integration: behavioral and electrophysiological validation. *Brain Topogr* 25:272–284.
- Kutas M, Federmeier KD (2011) Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annu Rev Psychol* 62:621–647.
- Kuznetsova A, Brockhoff PB, Christensen RH (2017) lmerTest package: tests in linear mixed effects models. *J Stat Soft* 82:1–26.
- Lenth RV (2022) emmeans: estimated marginal means, aka least-squares means. <https://github.com/rvlenth/emmeans>.
- Luck SJ (2014) Overview of common ERP components. In: An introduction to the event-related potential technique, pp 71–118. Cambridge, MA: Massachusetts Institute of Technology.
- Lupyan G (2015) Cognitive penetrability of perception in the age of prediction: predictive systems are penetrable systems. *Rev Phil Psychol* 6:547–569.
- Lupyan G, Abdel Rahman R, Boroditsky L, Clark A (2020) Effects of language on visual perception. *Trends Cogn Sci* 24:930–944.
- Machery E (2015) Cognitive penetrability: a no-progress report. In: The cognitive penetrability of perception: new philosophical perspectives (Zeimbekis J, Raftopoulos A, eds). Oxford, UK: Oxford UP.
- Maier M, Abdel Rahman R (2018) Native language promotes access to visual consciousness. *Psychol Sci* 29:1757–1772.
- Maier M, Abdel Rahman R (2019) No matter how: top-down effects of verbal and semantic category knowledge on early visual perception. *Cogn Affect Behav Neurosci* 19:859–876.
- Maier M, Glage P, Hohlfeld A, Abdel Rahman R (2014) Does the semantic content of verbal categories influence categorical perception? An ERP study. *Brain Cogn* 91:1–10.
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190.
- Matuschek H, Kliegl R, Vasishth S, Baayen H, Bates D (2017) Balancing Type I error and power in linear mixed models. *J Mem Lang* 94:305–315.
- Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh Inventory. *Neuropsychologia* 9:97–113.
- Palmeri TJ, Tarr MJ (2008) Visual object perception and long-term memory. In: Visual memory. (Luck SJ, Hollingworth A, eds), pp 163–208. Oxford, UK: Oxford UP.
- Panichello MF, Cheung OS, Bar M (2013) Predictive feedback and conscious visual experience. *Front Psychol* 3:620.
- Peikert A, Brandmaier AM (2021) A reproducible data analysis workflow with R Markdown, git, Make, and Docker. *QCMB* 1:e3763.
- Phelps EA, Ling S, Carrasco M (2006) Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychol Sci* 17:292–299.
- Pratt H (2011) Sensory ERP components. In: The Oxford handbook of event-related potential components. (Kappenman ES, Luck SJ, eds), pp 89–114. Oxford, UK: Oxford UP.
- Plylyshyn Z (1999) Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behav Brain Sci* 22:341–365.
- R Core Team R (2022) R: a language and environment for statistical computing. Vienna: Austria. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rossion B, Gauthier I, Goffaux V, Tarr MJ, Crommelinck M (2002) Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychol Sci* 13:250–257.
- Sagiv N, Bentin S (2001) Structural encoding of human and schematic faces: holistic and part-based processes. *J Cogn Neurosci* 13:937–951.
- Samaha J, Boutonnet B, Postle BR, Lupyan G (2018) Effects of meaningfulness on perception: alpha-band oscillations carry perceptual expectations and influence early visual responses. *Sci Rep* 8:6606.
- Sassenhagen J, Draschkow D (2019) Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology* 56:e13335.
- Schad DJ, Vasishth S, Hohenstein S, Kliegl R (2020) How to capitalize on a priori contrasts in linear (mixed) models: a tutorial. *J Mem Lang* 110:104038.
- Slivac K, Hervais-Adelman A, Hagoort P, Flecken M (2021) Linguistic labels cue biological motion perception and misperception. *Sci Rep* 11:17239.
- Supp GG, Schlögl A, Fiebach CJ, Gunter TC, Vigliocco G, Pfurtscheller G, Petsche H (2005) Semantic memory retrieval: cortical couplings in object recognition in the N400 window. *Eur J Neurosci* 21:1139–1143.
- Tanaka JW, Curran T (2001) A neural basis for expert object recognition. *Psychol Sci* 12:43–47.
- Teufel C, Dakin SC, Fletcher PC (2018) Prior object-knowledge sharpens properties of early visual feature-detectors. *Sci Rep* 8:10853.
- Teufel C, Nanay B (2017) How to (and how not to) think about top-down influences on visual perception. *Conscious Cogn* 47:17–25.
- Thierry G (2016) Neurolinguistic relativity: how language flexes human perception and cognition. *Lang Learn* 66:690–713.
- Thierry G, Martin CD, Downing PE, Pegna AJ (2007a) Is the N170 sensitive to the human face or to several intertwined perceptual and conceptual factors? *Nat Neurosci* 10:802–803.

4906 • J. Neurosci., June 28, 2023 • 43(26):4896–4906

Enge et al. • Semantic Information on Visual Perception

- Thierry G, Martin CD, Downing PE, Pegna AJ (2007b) Controlling for inter-stimulus perceptual variance abolishes N170 face selectivity. *Nat Neurosci* 10:505–511.
- Van Rossum G, Drake FL (2009) Python 3 reference manual. Scotts Valley, CA: CreateSpace.
- Voeten CC (2022) buildmer: stepwise elimination and term reordering for mixed-effects regression. <https://github.com/cvoeten/buildmer>.
- Weller PD, Rabovsky M, Abdel Rahman R (2019) Semantic knowledge enhances conscious awareness of visual objects. *J Cogn Neurosci* 31:1216–1226.
- Wickham H (2016) ggplot2: elegant graphics for data analysis. New York: Springer. <https://ggplot2.tidyverse.org>.
- Wickham H, et al. (2019) Welcome to the tidyverse. *J Syst Softw* 4: 1686.
- Wilke CO (2020) cowplot: streamlined plot theme and plot annotations for ggplot2. <https://wilkelab.org/cowplot/>.
- Wyatte D, Jilk DJ, O'Reilly RC (2014) Early recurrent feedback facilitates visual object recognition under challenging conditions. *Front Psychol* 5:674.
- Xu F, Tenenbaum JB (2007) Word learning as Bayesian inference. *Psychol Rev* 114:245–272.
- Yuille A, Kersten D (2006) Vision as Bayesian inference: analysis by synthesis? *Trends Cogn Sci* 10:301–308.

Instant Effects of Semantic Information on Visual Perception
Online Supplementary Information

Alexander Enge, Franziska Süß, & Rasha Abdel Rahman

Online Table 1. Unfamiliar Object Stimuli

Stimulus	ID	Matching keywords [German, English]	Non-matching keywords [German, English]
	1	elektrische Spannung, prüfen [electric current, measuring]	Makkaroni, formen [macaroni, forming]
	2	Makkaroni, formen [macaroni, forming]	Kuh, vom Zaun abhalten [cow, keeping away from the fence]
	3	Knochen, sägen [bones, sawing]	Streckenmaß, Sonnenlicht nutzen [distance measure, using sunrays]
	4	Unkraut, jäten [weed, removing]	Uhr, mit Wärme betreiben [clock, operating with heat]
	5	Mausefalle, zuschnappen [mousetrap, snap-shutting]	Brillenglas, zuschneiden [eyeglass lens, cutting]
	6	Goldmünzen, wiegen [gold coin, weighing]	Farbe, vom Fenster abschleifen [paint, scraping off the window]
	7	Knie, fixieren [knee, fixating]	Buchstaben, tippen [letter, typing]
	8	Eierkarton, pressen [egg carton, pressing]	Tabak, zermahlen [tobacco, pulverize]
	9	Baum, erklettern [tree, climbing]	Rotation, Ladung erzeugen [rotation, creating electric charge]
	10	Buchstaben, tippen [letter, typing]	Pferdehuf, Halt geben [horse hoof, giving grip]
	11	Akkordeon, spielen [accordion, playing]	Seil, schneiden [rope, cutting]

Table 1 continued

Stimulus	ID	Matching keywords	Non-matching keywords
	12	Körper, trainieren [body, training]	Buch, offen halten [book, binding]
	13	Farbe, vom Fenster abschleifen [paint, scraping off the window]	von Hand, zentrifugieren [by hand, centrifugating]
	14	Außerbereich, heizen [outdoor area, heating]	Rasierklinge, schärfen [razor blade, sharpening]
	15	Pflanzenteile, vergrößern [plant parts, magnifying]	Katzenklo, sich selbst reinigen [litter box, self-cleaning]
	16	Krawatten, aufhängen [necktie, hanging up]	Zeichnungen, vermessen [drawings, measuring]
	17	Schallplatte, abtasten [vinyl record, reading]	Ball, katapultieren [ball, catapulting]
	18	Glas, schneiden [glass, cutting]	Eier, wiegen [eggs, weighing]
	19	Briketts, pressen [briquette, pressing]	Narkosemittel, abgeben [anesthetic, administering]
	20	Orgelton, erzeugen [organ sound, making]	Bandage, rollen [bandage, rolling]
	21	Spannung, erzeugen [electric current, making]	Fußstütze, reiten [footrest, horseback riding]
	22	Narkosemittel, abgeben [anesthetic, administering]	Nüsse, aufbrechen [nuts, cracking]
	23	Bandage, rollen [bandage, rolling]	Schnee, rodeln [snow, sleighing]
	24	Weinfass, Loch einschlagen [wine barrel, smashing a hole]	Tier, einfangen [animal, trapping]

Table 1 continued

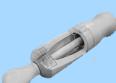
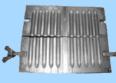
Stimulus	ID	Matching keywords	Non-matching keywords
	25	Flaschenkorken, einführen [bottle cork, inserting]	Pferd, im Moor laufen [horse, walking in the moor]
	26	Automat, Eier ausbrüten [machine, incubating eggs]	Windgeschwindigkeit, messen [wind speed, measuring]
	27	Korngarbe, greifen [sheaf, grabbing]	Treibhaus, heizen [glass house, heating]
	28	Brief, wiegen [letter, weighing]	Korngarbe, greifen [sheaf, grabbing]
	29	Löcher, bohren [hole, drilling]	Sonnenlicht, Intensität messen [sunlight, measure intensity]
	30	Angelschnur, kurbeln [fishing line, winding]	Glaskörper, musizieren [glass body, making music]
	31	Kurven, malen [curve, drawing]	Nussöl, pressen [nut oil, pressing]
	32	Radiofrequenz, einstellen [radio frequency, tuning]	Erektion, helfen [erection, helping]
	33	Zäpfchen, pressen [suppository, pressing]	Unkraut, jäten [weed, removing]
	34	Fass, öffnen [barrel, opening]	Pflanzenteile, vergrößern [plant parts, magnifying]
	35	Bergbaustollen, beleuchten [mining tunnel, lightening]	Knie, fixieren [knee, fixating]
	36	Kuh, vom Zaun abhalten [cow, keeping away from the fence]	Radiofrequenz, einstellen [radio frequency, tuning]
	37	Saatgut, gleichmäßig aussäen [seeds, sowing evenly]	Fass, öffnen [barrel, opening]

Table 1 continued

Stimulus	ID	Matching keywords	Non-matching keywords
	38	Flaschen, trocknen [bottles, drying]	Bleistift, anspitzen [pencil, sharpening]
	39	Uhr, mit Wärme betreiben [clock, operating with heat]	Körper, untersuchen [body, training]
	40	Tonpott, trommeln [clay pot, drumming]	Piano, stimmen [piano, tuning]
	41	Sprengstoffexplosion, auslösen [dynamite explosion, triggering]	Fisch, wiegen [fish, weighing]
	42	Pferdehuf, Halt geben [horse hoof, giving grip]	Zäpfchen, pressen [suppository, pressing]
	43	Bleistift, anspitzen [pencil, sharpening]	Angel, Köder markieren [fishing rod, marking bait]
	44	Nussöl, pressen [nut oil, pressing]	Zeichen, einbrennen [marks, burning in]
	45	Rasierklinge, schärfen [razor blade, sharpening]	Kurven, malen [curve, drawing]
	46	heiße Platten, anheben [hot plates, lifting]	Kleidung, im Eimer waschen [clothes, washing in a bucket]
	47	Film, aufspulen [film roll, winding]	Mund, offen halten [mouth, keeping open]
	48	Waffe, entflammen [weapon, inflaming]	Uhrzeit, anzeigen [time, displaying]
	49	Mikroskop-Proben, schneiden [microscopic samples, slicing]	heiße Platten, anheben [hot plates, lifting]
	50	Glaskörper, musizieren [glass body, making music]	Kork, flach pressen [cork, pressing flat]

SEMANTICALLY INFORMED PERCEPTION

ONLINE SI 5

Table 1 continued

Stimulus	ID	Matching keywords	Non-matching keywords
	51	Dampf, zerstäuben [steam, spraying]	Buch, binden [book, opening]
	52	Mandeln, operieren [kidneys, operating]	Film, aufspulen [film roll, winding]
	53	Toastbrot, rösten [toast, roasting]	Autobatterie, Spannung testen [car battery, measuring voltage]
	54	Fußstütze, reiten [footrest, horseback riding]	Stromstärke, messen [current, measuring]
	55	Türgelenk, Feuer überstehen [door hinge, surviving fires]	Botschaft, telegrafieren [message, telegraphing]
	56	Schnee, rodeln [snow, sledding]	Dampf, zerstäuben [steam, spraying]
	57	Nüsse, aufbrechen [nuts, cracking]	Ziegelsteine, formen [bricks, forming]
	58	Radiergummi, mit Strom betreiben [eraser, operating with electricity]	Münzen, aufbewahren [coins, storing]
	59	Treibhaus, heizen [glass house, heating]	Messer, schleifen [knife, sharpening]
	60	Botschaft, telegrafieren [message, telegraphing]	Toastbrot, rösten [toast, roasting]
	61	Angel, Köder markieren [fishing rod, marking bait]	Baum, erklettern [tree, climbing]
	62	Katzenklo, sich selbst reinigen [litter box, self-cleaning]	Tankfüllstand, messen [fuel tank level, gauging]
	63	Körper, untersuchen [body, examining]	Schuhe, auf Eis laufen [shoes, walking on ice]

Table 1 continued

Stimulus	ID	Matching keywords	Non-matching keywords
	64	Brillenglas, zuschneiden [eyeglass lens, cutting]	Saatgut, gleichmäßig aussäen [seeds, sowing evenly]
	65	Draht, wickeln [wire, winding]	Bergbaustollen, beleuchten [mining tunnel, lightening]
	66	Buch, offen halten [book, opening]	Sprengstoffexplosion, auslösen [dynamite explosion, triggering]
	67	Feldanbau, häckseln [harvest, chopping]	Draht, wickeln [wire, winding]
	68	Schnurlot, absenken [plummet, letting down]	Korken, formen [bottle cork, forming]
	69	Sternenbilder, vermessen [stellar constellation, measuring]	Schnurlot, absenken [plummet, letting down]
	70	Nachricht, morsen [message, signaling]	Feldanbau, häckseln [harvest, chopping]
	71	Musikgerät, stampfen [musical instrument, stamping]	Goldmünzen, wiegen [gold coin, weighing]
	72	Schuhe, auf Eis laufen [shoes, walking on ice]	Geschwindigkeit, ermitteln [speed, determining]
	73	Geschwindigkeit, ermitteln [speed, determining]	Mausefalle, zuschnappen [mousetrap, snap-shutting]
	74	Zeichen, einbrennen [marks, burning in]	Luftdruck, messen [air pressure, measuring]
	75	Tabak, zermahlen [tobacco, pulverize]	Flaschen, trocknen [bottles, drying]
	76	Stromstärke, messen [current, measuring]	Mandeln, operieren [kidneys, operating]

Table 1 continued

Stimulus	ID	Matching keywords	Non-matching keywords
	77	Tier, einfangen [animal, trapping]	Maiskolben, entkörnen [corn cob, removing grains]
	78	Messer, schleifen [knife, sharpening]	Schritte, vergrößern [steps, extending]
	79	Leierkasten, klingen [barrel organ, making sounds]	Fass, anheben [barrel, lifting]
	80	Buch, binden [book, binding]	Löcher, bohren [hole, drilling]
	81	Kork, flach pressen [cork, pressing flat]	Elektroschock, spielen [electric shock, playing]
	82	Tabletten, zerteilen [pill, splitting]	Blumentopf, sich selbst wässern [flowerpot, self-watering]
	83	Fass, anheben [barrel, lifting]	Mikroskop-Proben, schneiden [microscopic samples, slicing]
	84	Pferd, im Moor laufen [horse, walking in the moor]	Luft, abpumpen [air, pumping out]
	85	altes Ritual, hacken [ancient ritual, chopping]	Spannung, erzeugen [electric current, making]
	86	Kleidung, im Eimer waschen [clothes, washing in a bucket]	Türgelenk, Feuer überstehen [door hinge, surviving fires]
	87	Maiskolben, entkörnen [corn cob, removing grains]	Brief, wiegen [letter, weighing]
	88	Ball, katapultieren [ball, catapulting]	Radiergummi, mit Strom betreiben [eraser, operating with electricity]
	89	Zeichnungen, vermessen [drawings, measuring]	Brenner, löten [burner, soldering]

Table 1 continued

Stimulus	ID	Matching keywords	Non-matching keywords
	90	Elektroschock, spielen [electric shock, playing]	Weinfass, Loch einschlagen [wine barrel, smashing a hole]
	91	Lieferungen, abzählen [deliveries, counting]	Außenbereich, heizen [outdoor area, heating]
	92	Rotation, Ladung erzeugen [rotation, creating electric charge]	Musikgerät, stampfen [musical instrument, stamping]
	93	Herz, durch Maschine ersetzen [heart, replacing with machine]	Kerzen, löschen [candles, extinguishing]
	94	Seil, schneiden [rope, cutting]	Akkordeon, spielen [accordion, playing]
	95	Fisch, wiegen [fish, weighing]	Herz, durch Maschine ersetzen [heart, replacing with machine]
	96	Kerzen, löschen [candles, extinguishing]	Körper, trainieren [body, examining]
	97	Korken, formen [bottle cork, forming]	Sternenbilder, vermessen [stellar constellation, measuring]
	98	Erektion, helfen [erection, helping]	Waffe, werfen [weapon, throwing]
	99	Streckenmaß, Sonnenlicht nutzen [distance measure, using sunrays]	Kartoffeln, stampfen [potatoes, mashing]
	100	Luftdruck, messen [air pressure, measuring]	Knochen, sägen [bones, sawing]
	101	Piano, stimmen [piano, tuning]	Eierkarton, pressen [egg carton, pressing]
	102	von Hand, zentrifugieren [by hand, centrifugating]	Lieferungen, abzählen [deliveries, counting]

SEMANTICALLY INFORMED PERCEPTION

ONLINE SI 9

Table 1 continued

Stimulus	ID	Matching keywords	Non-matching keywords
	103	Kartoffeln, stampfen [potatoes, mashing]	Nachricht, morse [message, signaling]
	104	Waffe, werfen [weapon, throwing]	elektrische Spannung, prüfen [electric current, measuring]
	105	Tankfüllstand, messen [fuel tank level, gauging]	Tonpott, trommeln [clay pot, drumming]
	106	Uhrzeit, anzeigen [time, displaying]	altes Ritual, hacken [ancient ritual, chopping]
	107	Windgeschwindigkeit, messen [wind speed, measuring]	Waffe, entflammen [weapon, inflaming]
	108	Schlüsselloch, stanzen [keyhole, punching]	Automat, Eier ausbrüten [machine, incubating eggs]
	109	Ziegelsteine, formen [bricks, forming]	Schallplatte, abtasten [vinyl record, reading]
	110	Becher, Schall auffangen [drinking cup, picking up sound]	Schlüsselloch, stanzen [keyhole, punching]
	111	Luft, abpumpen [air, pumping out]	Briketts, pressen [briquette, pressing]
	112	Mund, offen halten [mouth, keeping open]	Orgelton, erzeugen [organ sound, making]
	113	Autobatterie, Spannung testen [car battery, measuring voltage]	Becher, Schall auffangen [drinking cup, picking up sound]
	114	Sonnenlicht, Intensität messen [sunlight, measure intensity]	Krawatten, aufhängen [necktie, hanging up]
	115	Kutschrad, anschließen [carriage wheel, locking]	Angelschnur, kurbeln [fishing line, winding]

Table 1 continued

Stimulus	ID	Matching keywords	Non-matching keywords
	116	Eier, wiegen [eggs, weighing]	Leierkasten, klingen [barrel organ, making sounds]
	117	Blumentopf, sich selbst wässern [flowerpot, self-watering]	Glas, schneiden [glass, cutting]
	118	Schritte, vergrößern [steps, extending]	Tabletten, zerteilen [pill, splitting]
	119	Münzen, aufbewahren [coins, storing]	Kutschrad, anschließen [carriage wheel, locking]
	120	Brenner, löten [burner, soldering]	Flaschenkorken einführen, einführen [bottle cork, inserting]

Online Results 1. Linear Mixed-Effects Models**P1 Component (100–150 ms)**

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: P1 ~ 1 + phase + condition + phase:condition + cosine + (1 +
##         phase | participant_id)
## Control: lme4::lmerControl
##
## REML criterion at convergence: 87306.7
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max 
## -6.4625 -0.6237 -0.0031  0.6134  5.8089 
##
## Random effects:
## Groups            Name        Variance Std.Dev. Corr
## participant_id (Intercept) 8.991    2.999
##                 phase1     1.670    1.292   -0.31
##                 phase2     1.822    1.350    0.44 -0.97
## Residual          25.582    5.058
## Number of obs: 14309, groups: participant_id, 48
##
## Fixed effects:
##             Estimate Std. Error       df t value Pr(>|t|)    
## (Intercept) 3.991e+00 4.386e-01 4.857e+01 9.099 4.55e-12 ***
## phase1     -6.827e-01 2.156e-01 4.797e+01 -3.167 0.00268 **  
## phase2      1.100e+00 2.230e-01 4.807e+01 4.930 1.02e-05 ***
## condition1 5.105e-02 1.054e-01 1.417e+04 0.484 0.62808    
## condition2 1.790e-03 1.219e-01 1.418e+04 0.015 0.98829    
## cosine      1.949e-01 7.413e-02 1.416e+04 2.629 0.00859 **  
## phase1:condition1 -2.392e-01 2.553e-01 1.350e+04 -0.937 0.34874    
## phase2:condition1  7.278e-01 2.563e-01 1.342e+04 2.840 0.00452 **
```

```

## phase1:condition2 -3.668e-01  2.933e-01  1.185e+04  -1.251  0.21113
## phase2:condition2  7.975e-01  2.946e-01  1.164e+04   2.707  0.00680 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) phase1 phase2 cndtn1 cndtn2 cosine phs1:1 phs2:1 phs1:2
## phase1     -0.265
## phase2      0.378 -0.857
## condition1  0.018  0.002  0.001
## condition2 -0.006  0.001  0.000  0.591
## cosine       0.127  0.000  0.001 -0.029 -0.042
## phs1:cndtn1  0.001  0.110 -0.053  0.002  0.002  0.002
## phs2:cndtn1  0.000 -0.055  0.110  0.006  0.005 -0.001 -0.511
## phs1:cndtn2  0.000 -0.002  0.002  0.002  0.002  0.000  0.582 -0.306
## phs2:cndtn2  0.000  0.002 -0.001  0.005  0.008  0.001 -0.306  0.584 -0.519
##
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## phase        730.61 365.30      2     49.7 14.2799 1.258e-05 ***
## condition     8.91   4.46      2 14172.1  0.1742  0.840161
## cosine        176.75 176.75      1 14159.0  6.9091  0.008585 **
## phase:condition 259.50  64.88      4 11197.7  2.5360  0.038129 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pairwise Contrasts (Simple Effects)
## contrast = Informed - Uninformed:
##           estimate   SE  df t.ratio p.value
## Pre-insight -0.0321 0.180 12943 -0.178  1.0000
## Insight      -0.2713 0.182 14138 -1.494  0.4060
## Post-insight  0.4565 0.181 13047  2.515  0.0357
##
## contrast = Informed - Unsuccessful:
##           estimate   SE  df t.ratio p.value
## Pre-insight -0.0195 0.206 10410 -0.095  1.0000
## Insight      -0.3863 0.209 13860 -1.844  0.1956
## Post-insight  0.4112 0.209 10673  1.972  0.1460
##
## Degrees-of-freedom method: satterthwaite
## P value adjustment: bonferroni method for 3 tests

```

N170 Component (150–200 ms)

```

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: N170 ~ 1 + phase + condition + phase:condition + cosine + (1 +
##           phase | participant_id)
## Control: lme4::lmerControl
##
## REML criterion at convergence: 86644.7
##
## Scaled residuals:
##       Min     1Q Median     3Q    Max
## -6.0470 -0.6299 -0.0035  0.6025  5.9459
##
## Random effects:
##   Groups      Name       Variance Std.Dev. Corr
##   participant_id (Intercept) 13.660   3.696
##           phase1      6.588   2.567     0.15
##           phase2      5.724   2.393     0.01 -0.96
##   Residual            24.250   4.924
## Number of obs: 14309, groups: participant_id, 48
##

```

```

## Fixed effects:
##                               Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)           1.810e+00  5.379e-01 4.801e+01   3.365 0.001511 **
## phase1              -3.057e+00  3.851e-01 4.751e+01  -7.938 2.93e-10 ***
## phase2               2.611e+00  3.612e-01 4.706e+01   7.229 3.65e-09 ***
## condition1          -1.670e-01  1.026e-01 1.417e+04  -1.627 0.103705
## condition2          -2.168e-01  1.187e-01 1.417e+04  -1.826 0.067843 .
## cosine                3.197e-01  7.218e-02 1.416e+04   4.430 9.51e-06 ***
## phase1:condition1  -5.275e-01  2.497e-01 1.413e+04  -2.113 0.034637 *
## phase2:condition1  9.032e-01  2.505e-01 1.402e+04   3.606 0.000313 ***
## phase1:condition2 -8.385e-01  2.879e-01 1.384e+04  -2.913 0.003586 **
## phase2:condition2  1.198e+00  2.888e-01 1.341e+04   4.148 3.37e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) phase1 phase2 cndtn1 cndtn2 cosine phs1:1 phs2:1 phs1:2
## phase1       0.143
## phase2      -0.922
## condition1  0.014  0.001
## condition2 -0.005  0.001  0.000  0.592
## cosine        0.101  0.000  0.000 -0.029 -0.042
## phs1:cndtn1  0.000  0.059 -0.032  0.003  0.003  0.002
## phs2:cndtn1  0.000 -0.030  0.065  0.005  0.005  0.000 -0.508
## phs1:cndtn2  0.000 -0.002  0.001  0.003  0.003  0.000  0.587 -0.304
## phs2:cndtn2  0.000  0.001 -0.001  0.005  0.007  0.001 -0.304  0.588 -0.514
##
## Type III Analysis of Variance Table with Satterthwaite's method
##                   Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## phase            1529.29  764.64     2     48.1 31.5319 1.784e-09 ***
## condition        92.04   46.02     2 14166.2  1.8976 0.1499606
## cosine           475.81  475.81     1 14157.3 19.6209 9.514e-06 ***
## phase:condition 489.34  122.34     4 13093.5  5.0448 0.0004633 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pairwise Contrasts (Simple Effects)
## contrast = Informed - Uninformed:
##   phase      estimate    SE   df t.ratio p.value
## Pre-insight -0.1163 0.176 13762 -0.662 1.0000
## Insight      -0.6439 0.177 14194 -3.631 0.0008
## Post-insight  0.2593 0.177 13742  1.462 0.4312
##
## contrast = Informed - Unsuccessful:
##   phase      estimate    SE   df t.ratio p.value
## Pre-insight -0.0571 0.202 12610 -0.282 1.0000
## Insight      -0.8956 0.205 14203 -4.371 <.0001
## Post-insight  0.3024 0.204 12556  1.479 0.4171
##
## Degrees-of-freedom method: satterthwaite
## P value adjustment: bonferroni method for 3 tests

```

N400 Component (400–700 ms)

```

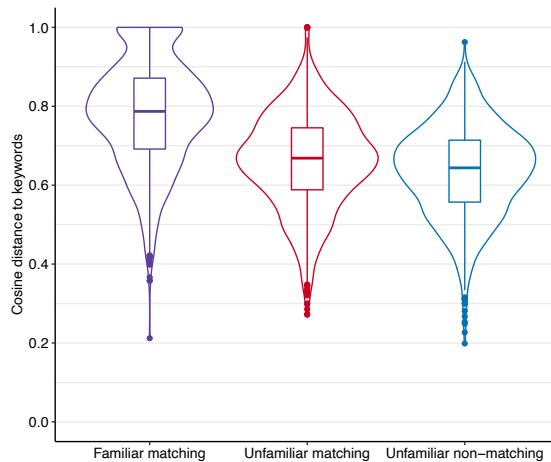
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: N400 ~ 1 + phase + condition + phase:condition + cosine + (1 +
##   phase | participant_id)
## Control: lme4::lmerControl
##
## REML criterion at convergence: 82101.6
##
## Scaled residuals:
##   Min     1Q Median     3Q    Max

```

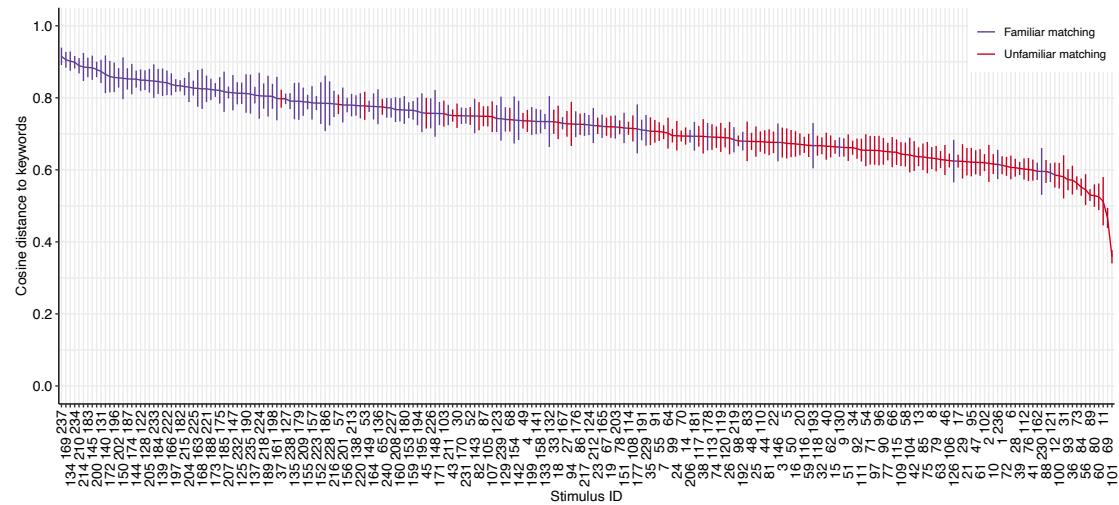
```

## -6.7184 -0.6289 -0.0010  0.6330  5.0726
##
## Random effects:
## Groups      Name      Variance Std.Dev. Corr
## participant_id (Intercept) 2.842   1.686
##                  phase1     1.566   1.251    0.19
##                  phase2     1.013   1.007   -0.03 -0.61
## Residual          17.754   4.214
## Number of obs: 14309, groups: participant_id, 48
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept) 1.145e+00 2.504e-01 5.053e+01 4.571 3.16e-05 ***
## phase1      2.628e+00 2.019e-01 4.746e+01 13.018 < 2e-16 ***
## phase2     -1.615e+00 1.712e-01 4.842e+01 -9.433 1.51e-12 ***
## condition1 5.837e-01 8.777e-02 1.418e+04 6.651 3.02e-11 ***
## condition2 5.025e-01 1.015e-01 1.420e+04 4.950 7.51e-07 ***
## cosine      5.312e-02 6.176e-02 1.416e+04 0.860 0.389735
## phase1:condition1 1.216e+00 2.135e-01 1.407e+04 5.694 1.27e-08 ***
## phase2:condition1 -2.903e-01 2.140e-01 1.383e+04 -1.356 0.175088
## phase1:condition2 1.317e+00 2.460e-01 1.362e+04 5.353 8.80e-08 ***
## phase2:condition2 -8.411e-01 2.465e-01 1.286e+04 -3.412 0.000647 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) phase1 phase2 cndtn1 cndtn2 cosine phs1:1 phs2:1 phs1:2
## phase1      0.162
## phase2     -0.027 -0.585
## condition1 0.027  0.002  0.001
## condition2 -0.009  0.001  0.000  0.591
## cosine      0.185  0.000  0.001 -0.029 -0.042
## phs1:cndtn1 0.001  0.097 -0.058  0.003  0.003  0.002
## phs2:cndtn1 0.000 -0.049  0.119  0.005  0.004  0.000 -0.500
## phs1:cndtn2 0.000 -0.003  0.001  0.003  0.003  0.000  0.586 -0.294
## phs2:cndtn2 0.000  0.001 -0.001  0.004  0.006  0.001 -0.294  0.587 -0.501
##
## Type III Analysis of Variance Table with Satterthwaite's method
##              Sum Sq Mean Sq NumDF DenDF F value    Pr(>F)
## phase        3097.77 1548.88      2     48.1 87.2427 < 2.2e-16 ***
## condition    813.58  406.79      2 14180.1 22.9129 1.162e-10 ***
## cosine       13.13   13.13      1 14159.6  0.7398  0.3897
## phase:condition 831.16 207.79      4 13624.4 11.7040 1.728e-09 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Pairwise Contrasts (Simple Effects)
## contrast = Informed - Uninformed:
## phase      estimate   SE  df t.ratio p.value
## Pre-insight -0.130 0.151 14117 -0.861 1.0000
## Insight      1.086 0.151 14156  7.173 <.0001
## Post-insight  0.795 0.152 14031  5.230 <.0001
##
## contrast = Informed - Unsuccessful:
## phase      estimate   SE  df t.ratio p.value
## Pre-insight -0.095 0.174 13794 -0.546 1.0000
## Insight      1.222 0.175 13925  6.998 <.0001
## Post-insight  0.381 0.176 13495  2.169 0.0903
##
## Degrees-of-freedom method: satterthwaite
## P value adjustment: bonferroni method for 3 tests

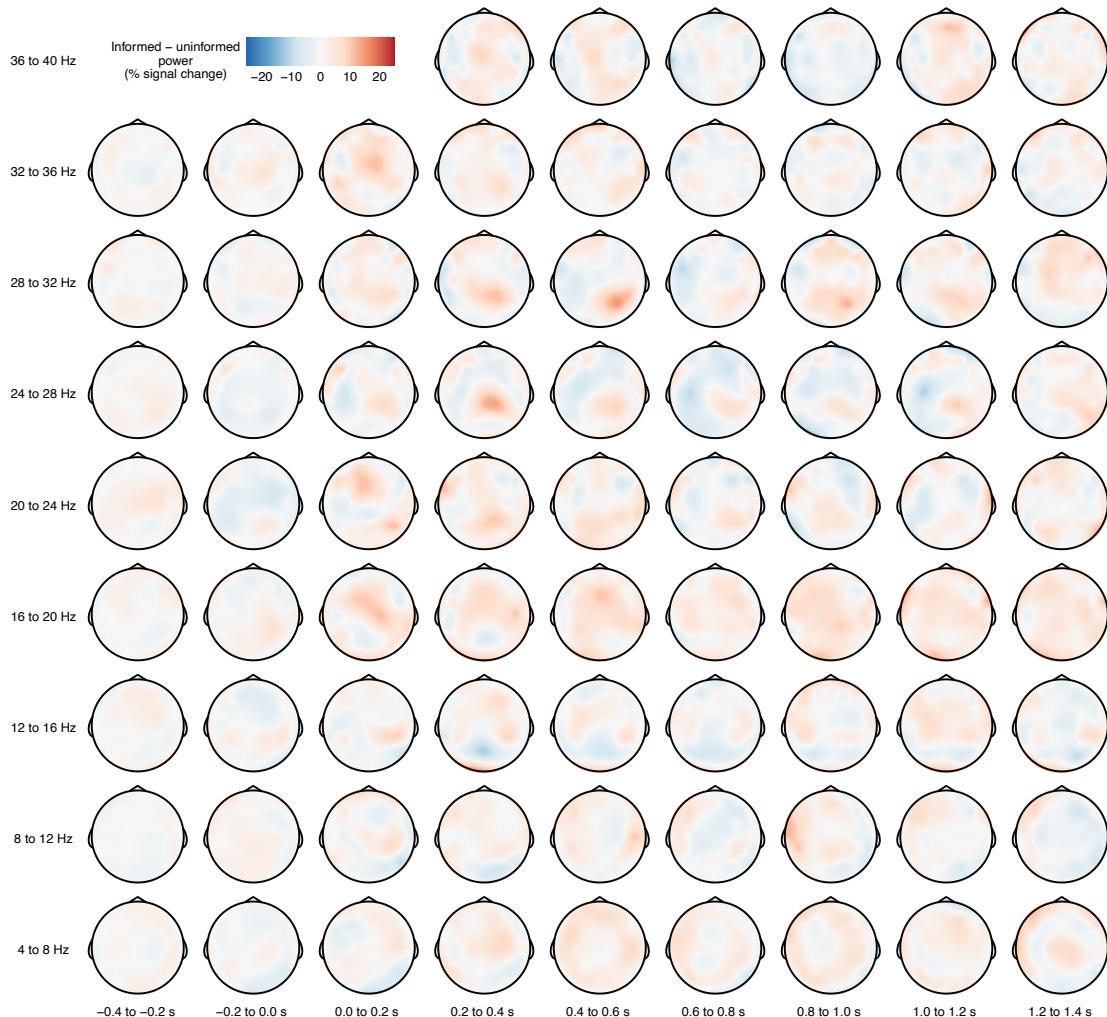
```



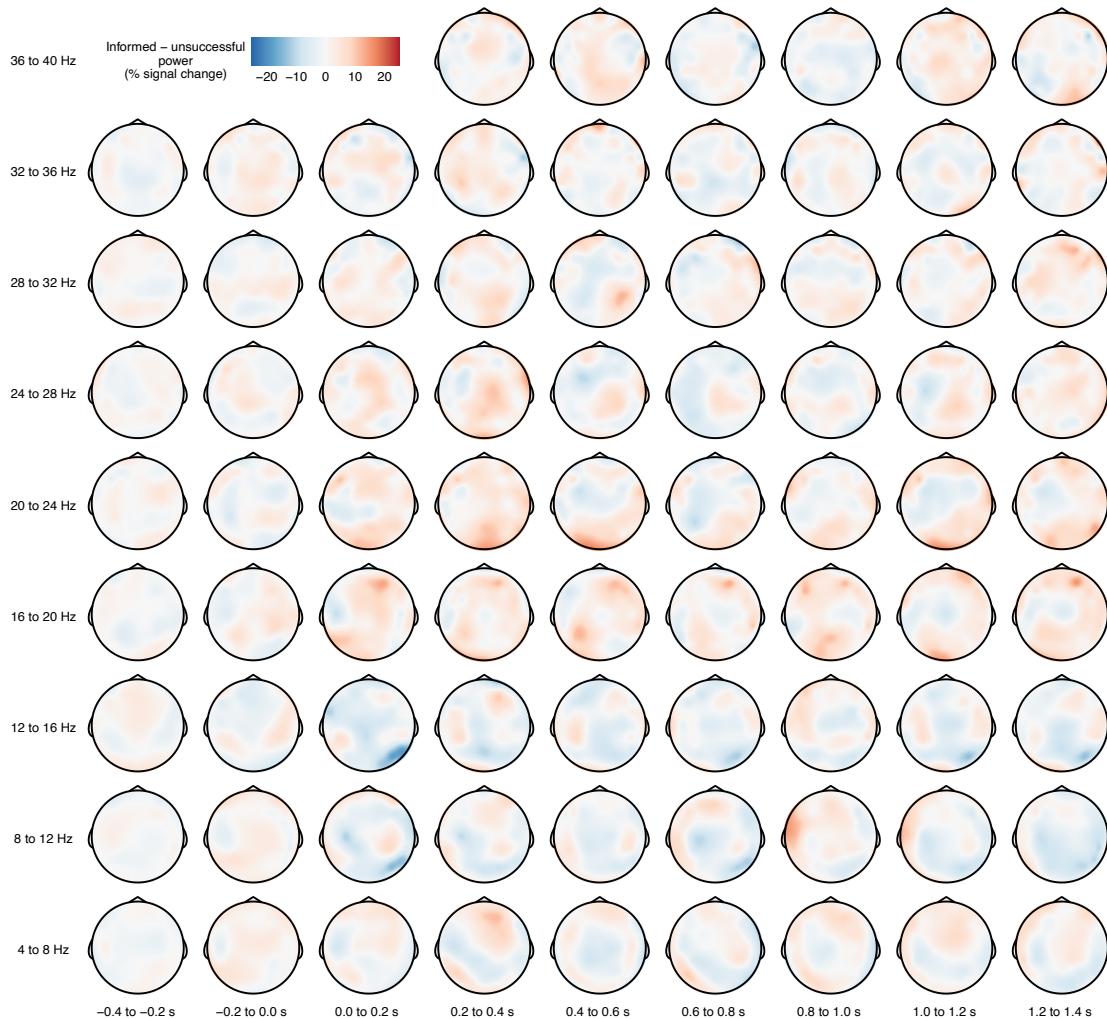
Online Figure 1. Online pre-rating study results. Participants were presented with 240 objects in random order, 120 of which were familiar everyday objects and the other 120 were presumed to be unfamiliar to most people. Participants were asked to describe or guess the function of each object by typing a pair of German keywords. Violins show the distributions of the similarities between these participant-generated keywords and the keywords that we had created for our main EEG experiment (see Materials and Methods; Online Table 1). We computed these similarities separately for (a) participant-generated keywords for the familiar objects and keywords that we had created to match the familiar objects (though these were not part of the main EEG experiment; red), (b) participant-generated keywords for the unfamiliar objects and keywords that we had created to match the unfamiliar objects (blue), and (c) participant-generated keywords for the unfamiliar objects and keywords that we had created to not match the unfamiliar objects (by selecting keywords that matched one of the other unfamiliar objects; purple). Semantic similarities were computed as the cosine similarity between the sums of the two word vectors in a word2vec embedding space pre-trained on the German Wikipedia. Boxplots show the median (thick line), 25th and 75th percentiles (hinges), 1.5 times the interquartile range above and below the hinges (whiskers), and any outlier data points that fall outside of the whiskers (dots).



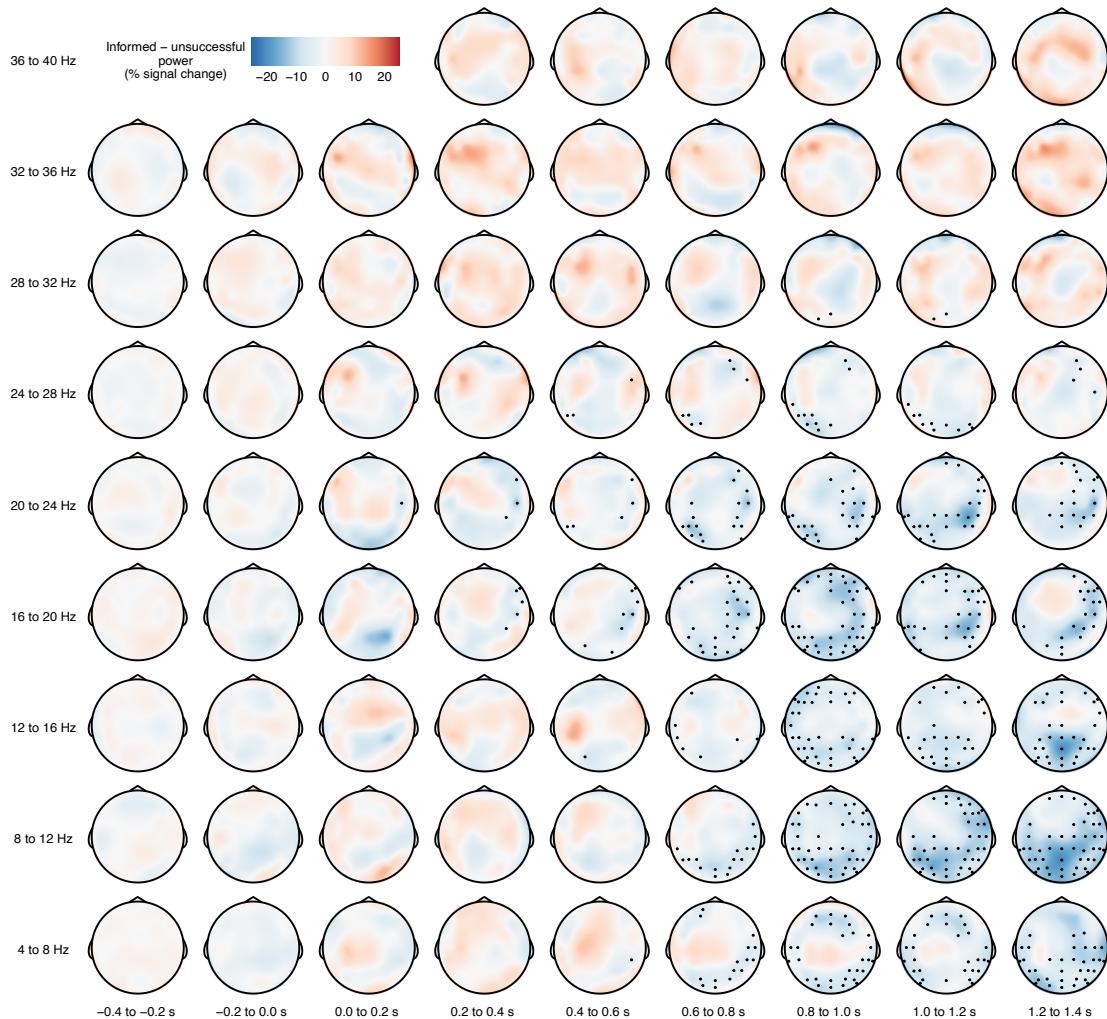
Online Figure 2. Online pre-rating study results on the level of individual object stimuli. Same as Online Figure 1, but showing, for each object stimulus separately, the mean (horizontal line) ± 1 standard error (vertical lines) of the semantic similarities between participant-generated keywords and the keywords that we had generated. Semantic similarities were generally higher for the familiar objects than for the unfamiliar objects, indicating that it was easier for participants to come up with the correct function for the familiar objects. This mean cosine similarity for each object (i.e., a measure of the difficulty of guessing its function) was entered as a covariate of no interest in all linear mixed-effects models for analyzing the data of the main EEG experiment (see Materials and Methods; Results).



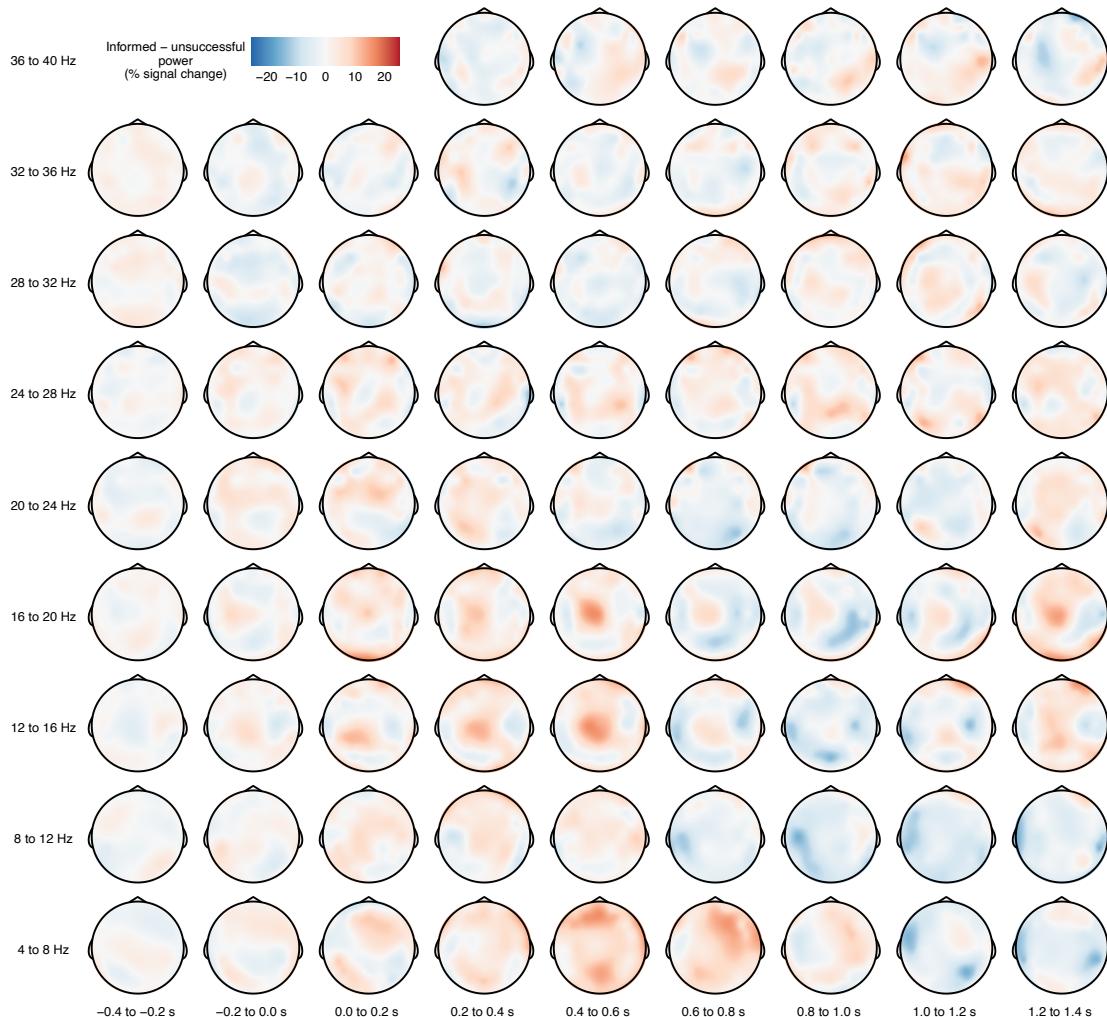
Online Figure 3. Time-frequency results for the pre-insight phase. Each topographic plot shows the difference in event-related power (in units of percent signal change) between the semantically informed condition and the uninformed condition, grand-averaged across participants. A cluster-based permutation test indicated no clusters for which this difference was statistically significant (all $p > .789$).



Online Figure 4. Time-frequency results for the pre-insight phase (informed minus unsuccessful). Each topographic plot shows the difference in event-related power (in units of percent signal change) between the semantically informed condition and the unsuccessfully informed condition, grand-averaged across participants. A cluster-based permutation test indicated no clusters for which this difference was statistically significant (all $ps > .770$).



Online Figure 5. Time-frequency results for the insight phase (informed minus unsuccessful). Each topographic plot shows the difference in event-related power (in units of percent signal change) between the semantically informed condition and the unsuccessfully informed condition, grand-averaged across participants. Black dots highlight EEG channels that were part of a cluster for which this difference was statistically significant ($p_{\text{cluster}} = .016$).



Online Figure 6. Time-frequency results for the post-insight phase (informed minus unsuccessful). Each topographic plot shows the difference in event-related power (in units of percent signal change) between the semantically informed condition and the unsuccessfully informed condition, grand-averaged across participants. A cluster-based permutation test indicated no clusters for which this difference was statistically significant (all $ps > .216$).

Acknowledgements

Thank you to my supervisors, Rasha Abdel Rahman and Michael Skeide, for encouraging me to pursue a doctoral degree and for your continued academic and emotional support. Your trust and guidance throughout my academic career have been a great honor, and you have helped me to learn not only a few little things about the human brain, but also many important things about myself. You are great mentors and I wish you continued success in your research and beyond.

Thank you to the members of the Research Group “Learning in Early Childhood” at the Max Planck Institute for Human Cognitive and Brain Sciences, especially to Roman Kessler and Anne-Sophie Kieslinger, and to the members of the Neurocognitive Psychology Lab at Humboldt-Universität zu Berlin, especially to Julia Baum, Anna Eiserbeck, Martin Maier, and Kirsten Stark. Your support and our countless interactions have made my PhD journey so much easier and so much more enjoyable. Should you decide to continue on this path, cognitive neuroscience has a bright future ahead with scientists like you.

Thank you to all participants who dedicated their time and energy to take part in my research studies as well as to all collaborators, research assistants, non-scientific staff, and funding agencies who have enabled this research.

Thank you to Roman Kessler for providing tremendously helpful feedback on my thesis, to Rasha Abdel Rahman, Gesa Hartwigsen, Sebastian Marktett, Gesa Schaad, and Peter Weller for serving on my dissertation committee, and to Rasha Abdel Rahman, Gesa Hartwigsen, and Gesa Schaad for their written reviews.

But most of all, thank you to my family and friends, especially to my parents and to Gabriele, Julia, Konstantin, Lisa, and Nele. Without your love, none of this would have any meaning.

Selbständigkeitserklärung

Hiermit erkläre ich,

- dass keine Zusammenarbeit mit gewerblichen Promotionsberatern stattfand,
- dass ich die dem angestrebten Verfahren zugrundeliegende Promotionsordnung zur Kenntnis genommen habe,
- dass die Dissertation oder Teile davon nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht, angenommen oder abgelehnt wurden,
- dass ich mich nicht anderwärts um einen Doktorgrad beworben habe bzw. einen entsprechenden Doktorgrad besitze,
- dass die Dissertation auf der Grundlage der angegebenen Hilfsmittel und Hilfen selbstständig angefertigt worden ist und
- dass die Grundsätze der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

Alexander Enge

Berlin, den 27. April 2025

