THE ENLANGUAGED BRAIN

Cognitive and Neural Mechanisms of Linguistic Influence on Perception

Ksenija Slivac

International Max Planck Research School (IMPRS) for Language Sciences
The educational component of the doctoral training was provided by the International Max Planck Research School (IMPRS) for Language Sciences. The graduate school is a joint initiative between the Max Planck Institute for Psycholinguistics and two partner institutes at Radboud University – the Centre for Language Studies, and the Donders Institute for Brain, Cognition and Behaviour. The IMPRS curriculum, which is funded by the Max Planck Society for the Advancement of Science, ensures that each member receives interdisciplinary training in the language sciences and develops a well-rounded skill set in preparation for fulfilling careers in academia and beyond. More information can be found at www.mpi.nl/imprs

The MPI series in Psycholinguistics
Initiated in 1997, the MPI series in Psycholinguistics contains doctoral theses produced at the Max Planck Institute for Psycholinguistics. Since 2013, it includes theses produced by members of the IMPRS for Language Sciences. The current listing is available at www.mpi.nl/mpi-series

ISBN: 978-94-92910-39-4
Cover: Julia Misersky
Layout: Ksenija Slivac
Printed and bound by Ipskamp Drukkers, Enschede

The research reported in this thesis was conducted at the Max Planck Institute for Psycholinguistics, in Nijmegen, the Netherlands

The Enlanguaged Brain:

Cognitive and Neural Mechanisms of Linguistic Influence on Perception

Proefschrift

ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de rector magnificus prof. dr. J.H.J.M. van Krieken,

volgens besluit van het college voor promoties

in het openbaar te verdedigen op

maandag 27 juni 2022

om 11.30 uur precies

door

Ksenija Slivac

geboren op 2 juli 1986

te Virovitica, Kroatië

Promotor:

Prof. dr. Peter Hagoort

Copromotor:

Dr. Monique E.P. Flecken (Universiteit van Amsterdam)

Manuscriptcommissie:

Prof. dr. Floris P. de Lange

Prof. dr. Gary Lupyan (University of Wisconsin-Madison, Verenigde Staten)

Prof. dr. Rasha Abdel Rahman (Humboldt-Universität zu Berlin, Duitslan

TABLE OF CONTENTS

CHAPTER 1

General Introduction

Back in the 17th century, Galileo Galilei warned us against the deceptive appearances of the world. Much to the disbelief of his contemporaries, he claimed that our senses do not accurately reflect physical reality, but are merely the product of our consciousness. While his claims seemed controversial at the time, he was not wrong. In fact, since then, as our understanding of the human brain progressed, we have learnt that our perception of seemingly objective reality – the way we view and experience the world – is far from objective.

When receiving inputs from our surroundings, we are not only constrained by biological affordances and mechanisms of our sensory systems and their encoding abilities, but also by a myriad of preconceived notions. This is not necessarily a bad thing from an evolutionary perspective. In fact, it is a very useful and computationally efficient property. Our perception has evolved to take physical data from our surroundings and turn it into meaningful, useful percepts based on our prior knowledge, expectations, attentional resources, and other cognitive processes. In other words, we see the behavioural value of the data surrounding us, rather than the data itself in its raw format. As useful as this practice is, it also highlights the fact that we see the world around us through the lens of conventionality and behavioural utility. We have become quite adept at not only isolating the useful and meaningful in any given context, but also at silencing the irrelevant or out-of-context input (Gilbert & Li, 2013; Gilbert & Sigman, 2007). Several studies have shown the remarkable extent to which people can miss the presence of a clearly visible, yet out-of-context stimulus in a noisy environment, as a function of attentional top-down influence on perception (Simons & Chabris, 1999; Drew et al., 2013).

Today, with a growing number of neuroscientific studies investigating the interaction between cognitive and perceptual systems, the question is no longer whether cognitive processes can affect perception, but rather to what extent that happens at any given moment and how strongly they dictate what we do and do not see. Moreover, with the rise of neuroimaging methods such as functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) and electroencephalography (EEG), we can now look at the neural correlates of such effects and investigate the cognitive and neural mechanisms by which these top-down influences arise.

1.1. The Power of Language

Consistent with the view that our perception is constantly shaped by cognitive and contextual influences, Edward Sapir and Benjamin Lee Whorf have argued that language exerts a particularly strong influence on thoughts and actions. This supposition, today known as the Sapir-Whorf hypothesis, has started a line of thinking and research still not resolved to this day. The radical interpretation of this hypothesis, called linguistic determinism, has argued that the language we speak determines how we think about and the extent to which we *can* understand the world. The more moderate interpretation, linguistic relativism, sees the linguistic influence on thought and perception more as a nuanced and context-flexible bias, rather than a set of constraints. While both interpretations of the Sapir-Whorf hypothesis were initially quickly dismissed by linguists and psychologists alike, the more moderate interpretation was brought back to life in the past several decades, with the rise of modern cognitive neuroscience and theories on grounded cognition (Barsalou, 1999; Casasanto, 2012; Gumperz & Levinson, 1991).

The grounded cognition account posits that high-level cognitive processes are closely intertwined with sensory, motor, and affective systems. Rather than seeing all these systems as separate, impenetrable modules in the brain, grounded cognition considers them to be mutually influential and highly interactive (Barsalou, 1999, 2008). Although accepted in cognitive neuroscience at large, this line of thinking has had a particularly strong impact on our understanding of language processing. Namely, the grounded cognition approach argues that our cognitive activities, and by extension language, are rooted in our sensory-motor processes and situated in specific contexts, i.e., they reactivate the same neural activation patterns present when we perceive or interact with objects or situations they refer to (Barsalou, 2008, 2009; Borghi & Cimatti, 2009). In other words, we understand language in part by means of recruiting the same neurons activated when we actually interact with what has been expressed through language.

A number of studies have shown that language, much like perception, is not a neurally isolated system solely contained within cortical perimeters traditionally assigned to it, i.e., within the left temporal and inferior frontal regions in the brain (Binder et al., 2009; Hoenig et al., 2008; Huth et al., 2016). Instead, linguistic processes can engage a widely distributed network in the brain in a flexible, dynamic and context-dependent manner. Moreover, research has shown that language can spontaneously engage perceptual regions, and can even modify how we receive and interpret visual inputs. For example, a number of studies have found that language can affect the perception of visual stimuli such as colour (Forder et al., 2016; Forder & Lupyan, 2019), shapes and objects (Stanfield & Zwaan, 2001; Zwaan et al., 2002), size (de Koning et al., 2017), contrast (Pelekanos & Moutoussis, 2011), and different types of motion (Bidet-Ildei et al., 2011; Dils & Boroditsky, 2013; Francken, Kok, et al., 2015; Meteyard et al., 2007).

Further studies have argued that linguistic labels might have a somewhat special status among higher-level cognitive systems in their ability to engage and modify perceptual processes. A particularly strong body of evidence for those claims comes from studies showing that language can speed up or potentiate our ability to perceive stimuli that are effectively invisible to the human eye and the brain – it can make the invisible visible (Boutonnet & Lupyan, 2015; Lupyan & Spivey, 2010a; Lupyan & Ward, 2013; Ostarek & Huettig, 2017).

The remarkable capacity of language to modify perception has been addressed in the label-feedback hypothesis (Lupyan, 2012; Lupyan et al., 2020). According to this proposal, linguistic ability to readily engage non-linguistic functional systems stems from the understanding that linguistic labels are unmotivated – untied to a particular instance or exemplar of a category. As such, they allow us to extract diagnostic features of a lexicalised concept and the category it represents without unnecessary details that pictures or sounds cannot escape. As a result, conceptual information delivered through labels is strong enough to co-engage perceptual representations of those features, thus warping or biasing our perceptual activation towards the labelled category (Lupyan, 2012; Lupyan et al., 2020). This claim builds upon more general accounts on the feature-based nature of top-down processes proposed by non-language studies, documenting the importance of categorical features for the detection of visual inputs (Ling et al., 2009; Maunsell & Treue, 2006; Saenz et al., 2002; Summerfield & Egner, 2016; Theeuwes, 2013).

While there are many behavioural studies reporting linguistic ability to affect perception, only a few neuroimaging studies have investigated neural patterns underlying such top-down influence. Several EEG studies have confirmed the label advantage narrative, by showing that language can modify perceptual processes as soon as 100-200ms after the incoming input (Boutonnet & Lupyan, 2015; Landau et al., 2010; Noorman et al., 2018). Findings from fMRI studies looking into the neural signatures of language-perception interaction, however, are much more divided (Francken, Kok, et al., 2015; Pirog Revill et al., 2008; Puri et al., 2009). The few that have been done on the topic have looked exclusively at changes in the overall magnitude of activation as proof of linguistic influence on perception. However, none of them looked at more subtle, differential patterns of activation across voxels within regions responding to perceptual representations of features expressed by the labels.

The focus of linguistic cueing studies on the (increased or decreased) overall magnitude of activation solely is an important detail to mention, especially because recent non-linguistic cueing studies have been focusing on an alternative, more subtle neural signature of top-down influences on perception. Namely, they have reported neural sharpening as the mechanism underlying the effect of top-down expectations on perception (Kok et al., 2012; Martens & Gruber, 2012; Yon et al., 2018; K. Zhang & Sejnowski, 1999). The sharpening account postulates that top-down expectations influence perceptual neural activation in such a way that the expected visual inputs are upregulated (i.e., enhanced), while the unexpected inputs are simultaneously downregulated (i.e., dampened). This shift in the pattern of activation can be decoded from the brain with multivariate pattern analysis (MVPA) approaches, but depending on the ratio of up- and downregulated neurons in a region of interest, it can go undetected when standard univariate approaches are employed. In this thesis, we therefore employ both univariate and multivariate approaches when examining neural signatures of linguistic top-down influences on perception.

1.2. Motion Perception

In examining lexical influence on perception, the experimental chapters of this thesis have focused on a particular type of perception: motion perception. The ability to perceive and quickly recognize movements around us is one of the most evolutionarily salient features of our visual system. Changes in our environment usually provide important information for our behaviour, which is reflected in our ability to detect and contextualise motion very quickly. We are particularly adept at recognising and categorising motion patterns, even in a relatively noisy environment. Generally speaking, we can distinguish between two motion categories: general (also referred to as global, rigid, or abstract) and biological motion perception. While general motion is usually associated with inanimate entities in our surroundings (such as rain or wind, for example), biological motion is identified as a property of animate, living beings. Even though, by its (dynamic) nature, motion perception forces us to rely on a number of exogenous cues that go beyond the motion itself, we can gather a surprising amount of data about both of these types of motion from the kinematic patterns alone (Beintema et al., 2006; Ramsey & Hamilton, 2010).

Relying on the human ability to quickly recognise and categorise kinematic patterns, researchers have used various types of point-light stimuli or displays (also known as point-light kinematograms; PLKs) to systematically test human ability to detect motion. These stimuli are compositional, in that they do not have a clear contour. Rather, the final percept they depict requires an integration of the constituent parts – point lights – into a coherent whole, based on the spatial configuration and kinematic patterns of those parts.

While these types of stimuli can be constructed in many ways, the most commonly employed point-light kinematograms, typically used in general motion perception studies, are random dot motions (RDMs; see **Figure 1.1.A**). This type of stimuli gives scientists the ability to easily manipulate motion parameters such as direction, speed, duration, density, etc. A typical RDM consists of a number of dots moving through space and time at a certain level of coherence. For example, for a 10% coherent motion display, 10% of the dots inside of an RDM aperture move in the same direction while the remaining 90% of the dots move randomly. Naturally, the higher the percentage of coherently moving dots, the easier it is to perceive the global motion direction within an RDM.

A special type of point-light kinematograms, knows as point-light figures (PLFs; see **Figure 1.1.B**), has emerged as the standard for studying biological motion perception (Johansson, 1973). Johansson (1973) pioneered the systematic investigation of biological motion perception, documenting a remarkable ease with which we can recognize and classify motion, based on a handful of spatially distributed dots with synchronised kinematics. Namely, he demonstrated that simply illuminating the joints of a human body and subsequently extracting those points into a dynamic array indicative of human motion suffices to project a vivid percept of biological motion. The process of perceiving these stimuli therefore requires a correct grouping of constituent dots and visual integration of their kinematics over space and time into a coherent percept of a human being in motion. Interestingly enough, the simple act of inverting a point-light figure or scrambling the starting points of the joints comprising the figure compromises the coherence of the final percept and renders it nonsensical (Bertenthal & Pinto, 1994; Pastukhov, 2017; Reed et al., 2003). This discovery made point-light displays the most commonly used means of testing both general and biological motion perception.

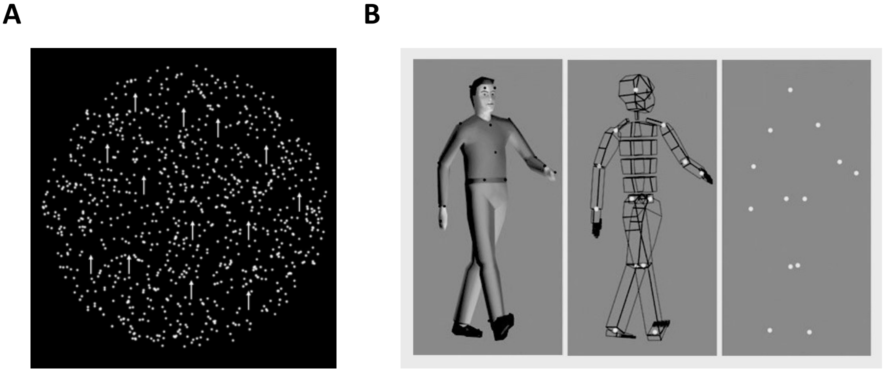


Figure 1.1. Point-light stimuli. (**A**) Random dot motion (RDM) with arrows indicating the dominant motion vector. (**B**) The process of point-light figure (PLF) creation: from the identification of major joints to the attachment of point lights to those joints and extraction of the final configuration. The figures are adapted from Slivac et al. (2021; figure A) and Vanrie & Verfaillie (2004; figure B)

A significant portion of the visual system is dedicated to motion perception, with the distinction between general and biological motion also present in neural encoding patterns therein. The posterior medial temporal (MT/V5) and medial superior temporal regions (MST) have been shown to respond robustly to more abstract, translational and rotational motion patterns (Born & Bradley, 2005; Kourtzi & Kanwisher, 2000; Pilly & Seitz, 2009; Poirier et al., 2005; Tootell et al., 1995). Conversely, biological motion perception has been shown to activate a complex network comprised of the lateral occipito-temporal cortex (LOTC), the posterior superior temporal sulcus (pSTS), the fusiform gyrus (FG), in particular the fusiform body area (FBA) therein, and the inferior parietal lobule (IPL), as well as the premotor cortex (Bonda et al., 1996; Gilaie-Dotan et al., 2013; Grèzes et al., 2001; Grossman et al., 2000; Grossman & Blake, 2002; Saygin, 2004, 2007a, 2007b; van Kemenade et al., 2012).

In this thesis, we make use of both the categorical and neural dichotomy between biological and general motion perception, to study cognitive and neural mechanisms of lexical top-down influences on perception. The compositional nature of point-light stimuli allowed us to manipulate presence and absence of cue-relevant features, and go a step further than previous research in investigating how language can affect visual integration processes. How we go about doing that across studies is presented in the overview of the thesis below.

1.3. Thesis Overview

This thesis addresses the question of linguistic influence on perception, i.e., the extent to which linguistic labels can influence how we see and interpret visual inputs, as well as the neural signatures underlying such influence. Specifically, studies reported in this thesis look at how lexical cues encoding biological or general motion features affect the perception of biological and general motion, shown through point-light stimuli in a dynamic and bistable environment. We employ behavioural, psychophysical, and neuroimaging (fMRI) approaches to examine both cognitive and neural processes by which these types of influences arise.

Chapter 2 provides an overview of the current state of knowledge regarding linguistic effects on perception, by summarising findings from literature investigating linguistic top-down influences on perception. In the review, we propose that the observed shifts in how we receive and interpret visual inputs as a function of linguistic top-down effects are the result of linguistic grounding. Finally, we relate the summarised findings to more generally observed mechanisms underlying non-linguistic top-down (expectation and attention) influences on perception.

Chapter 3 offers insight into the nature of language-perception interaction by empirically attesting feature activation as the driving mechanism behind lexical influence on perception. Employing dynamic, point-light stimuli with manipulable feature representation allowed us to go a step further than the existing cueing studies in showing that the linguistically mediated perceptual warping grows in strength as a function of cue-target feature accessibility and overlap. This influence is robust enough to override the visual input and mislead observers into constructing an erroneous percept when a cue is incongruent with the visual input, even when the visual input is highly detectable.

Chapter 4 investigates neural signatures underlying linguistic cueing effects on perception of two types of motion – biological and general. In particular, we report findings from two experiments – biological and general motion discrimination – testing if lexical cues can modify neural activation levels in biological and general motion regions, as well as early visual areas. Behaviourally, we found that lexical cues congruent with the visual target facilitate the perception of that target. Looking at the neural correlates of that facilitation, we show that lexical cues congruent with the target lead to an overall increase in the neural activation in perceptual regions encoding the visual representation of that target (the lateral occipito-temporal complex and the inferior parietal lobule for biological motion and the MT/V5 and early visual cortex for general motion perception, respectively).

Chapter 5 employs multivariate pattern analysis in order to investigate more subtle patterns of activation underlying lexical top-down influence on perception, in both higher- and lower-level perceptual regions. We use the same cues and targets as in Chapter 4, to test whether lexical cueing effects can be decoded from motion perception and early visual areas. We show that we can successfully decode shifts in activation patterns in both higher and lower perceptual regions as a function of lexical cues. Further, looking at changes in BOLD signal extracted from significant searchlight clusters, we show that the classifier distinguishing between target-congruent cues and no language cues is picking up on more uniform changes in BOLD signal within those clusters, also detectable with the univariate approaches. However, the classification between two different lexical cues relies on a more subtle interplay between activation suppression and enhancements, akin to the sharpening account proposed by studies looking at the effect of non-linguistic cues on perception. Looking further into the neural signatures of false alarms in the visual cortex, we find changes in the patterns of neural activation therein compared to correct rejections, suggesting that false alarms precede the decision-making level in the visual discrimination task.

Finally, in **Chapter 6,** we summarise the aforementioned chapters and discuss the significance of our findings for the understanding of cognitive and neural mechanisms underlying linguistic top-down influences on perception.

CHAPTER 2

Grounding the Label Advantage for Perception

Abstract

To date, research has reliably shown that language can engage and modulate perceptual processes in a top-down manner. However, our understanding of cognitive and neural mechanisms underlying such top-down influences is still under debate. In this review, we provide an overview of findings from literature on grounded cognition and linguistic cueing studies, in an effort to isolate such mechanisms. Additionally, we connect the findings from linguistic cueing studies to those reported in non-linguistic cueing literature, in order to find commonalities in neural processes allowing for top-down influences on perception. In doing so, we discuss the effects of language on perception in the context of broader, general cognitive and neural principles. Finally, we propose a way forward in the study of linguistic influences on perception.

In preparation as: Slivac K., Flecken M., Hagoort P.(in prep). Grounding the Label Advantage for Perception.

2.1. Introduction

The language system in the brain is extensively intertwined with other functional systems, such as perception, attention, emotions, and the sensory-motor system (Binder et al., 2009; Huth et al., 2016). In the last few decades, the extent to which language can and does regularly affect these systems, especially higher and lower level perception, and modify neural activation therein, has become of particular interest to cognitive (neuro)scientists (Barsalou, 2008; Borghi & Cimatti, 2009; Dove, 2020; Louwerse, 2011; Tillas, 2015).

The line of research concerned with the ability of language to engage perceptual processes is quite multifaceted. On the one hand, neuroscientists interested in neural correlates of linguistic processing have studied language in the context of the *grounded cognition* narrative. This account views cognition in general, and therefore also language, as interacting with our perceptual, motor and introspective systems, and being capable of stimulating those systems in order to re-enact experiences with our surroundings, body and mind (Barsalou, 2008). Building upon a grounded view of cognition, research focused on language acknowledges the fact that language allows us to systematically document, think, and communicate about the world around us, facilitating both retention and transmission of our thoughts and experiences. By forming associations between linguistic and non-linguistics aspects of our cognition and sensation, we end up using language as a shortcut or tool for conceptualisation, removing the necessity to fully experience first-hand everything that has been verbally expressed in order to understand it. In other words, with language we can simulate our experiential states in a retroactive manner, both behaviourally – by reinforcing or augmenting our understanding of what has been expressed through language, and neurally – by co-activating neural systems that encode such states (Barsalou, 2008; Borghi & Cimatti, 2009; Tillas, 2015).

On the other hand, neuroscientists interested in perception have investigated the extent to which the perceptual systems are susceptible to top-down influences such as attention, expectation, prior beliefs, and also language (Gilbert & Li, 2013; Gilbert & Sigman, 2007). Top-down influences on perception have been studied mainly by employing a cueing paradigm, with a cue, such as a symbol, word, picture, or sound, presented shortly before a visual stimulus (target) upon which participants are asked to perform a task. This approach allows for a highly controlled environment in which researchers can outline how the presence of a preceding stimulus can affect the perception of the visual target and/or the neural pattern of activation in response to that target.

This paradigm has also been adopted by researchers interested in linguistic top-down effects on perception. A series of behavioural and neuroimaging studies have revealed that language seem to be particularly efficient in modulating our perceptual experiences. In particular, several cueing studies have directly compared the effect of linguistic labels (e.g., word *dog*) to that of visual (e.g., picture of a dog) or auditory (e.g., barking sound) cues on visual perception, showing that people respond to a visual task more quickly and accurately when cued by linguistic rather than visual or auditory cues (Boutonnet & Lupyan, 2015; Lupyan & Spivey, 2010a). These findings have put forward the notion that linguistic labels exert a stronger influence on perceptual processes than seemingly equivalent environmental sounds or pictures (Edmiston & Lupyan, 2015).

Theoretical Account of Language-Perception Interaction

While these two lines of research – grounded cognition and linguistic (cueing) top-down influences on perception – may not seem closely connected in the literature, they converge on the mechanism by which language engages the perceptual system. And indeed, the peculiar ability of language to affect perceptual processes has been formalised as the label-feedback hypothesis by Lupyan et al. (2012) in a way that addresses and connects findings coming from both the grounded cognition literature and perception research. This account proposes that linguistic labels (i.e., words) are the perfect vehicle for categorization. By being unmotivated – untied to a particular exemplar - linguistic labels create a tool for conceptualisation that allows one to extract the diagnostic features of the lexicalised concept and the category it belongs to without the superfluous details (Edmiston & Lupyan, 2015). That lack of non-diagnostic details makes the category-diagnostic information particularly prominent and allows for the creation of a strong, stable and noise-free conceptual representation capable of engaging the perceptual system and biasing (‘warping’) its activation towards the labelled concept (for a detailed account of the LFH, see Lupyan, 2012; Lupyan et al., 2020).

The two main claims of the label-feedback hypothesis – that language modifies perception in a feature-based manner and that the key to its efficiency lies in the absence of evoking representations with unnecessary details – are not new in the field of neuroscience, although they are rarely referenced in neurolinguistic research. The feature-based nature of top-down influences on perception has been well documented in studies on the effect of attention and expectation on perception (Maunsell & Treue, 2006; Saenz et al., 2002; Summerfield & Egner, 2016). Namely, scientists have found that the representations of components or features of a visual input can be enhanced by means of top-down attentional and expectational influences. Building upon feature-based accounts of attention and expectation, the label-feedback hypothesis simply states that linguistic labels optimise the extraction of relevant (i.e., category-diagnostic) features (by not being tied to a specific exemplar), and in doing so end up being particularly efficient in feature-based top-down influence on perception. The detrimental effects of too many details, dubbed the dilution effect*,* has also been well documented in studies on decision making. The dilution effect describes the finding that our brains do not sum up, but rather average across pieces of information we receive. This means that the more pieces of information the brain receives, the less of a weight each individual piece of information has, thus diluting the information crucial for category construal. In other words, superfluous details, even when accurate, dilute the weight of category diagnostic features and end up creating a less strong conceptual (and perceptual) representation in the brain (Hotaling et al., 2015; Nisbett et al., 1981).

In sum, the current theoretical stance on linguistic effects on perception is that, by being untied to any particular category exemplar – and therefore optimally objective and restricted to defining categorical features only – linguistic labels can be grounded in perception in a particularly strong and noise-free manner, resulting in a stronger effect on any potential visual input. The basis of this theoretical postulation is in line with the (non-linguistic) top-down and information processing mechanisms reported in the cognitive neuroscience literature, which is an important step towards studying language in the context of general cognitive and neural mechanisms in the brain.

Review Overview

We have come a long way in understanding how language affects perception. However, the conditions necessary for linguistic effects on perception to be observed as well as the exact neural mechanism behind this type of effect are still under debate (Firestone & Scholl, 2016; Pylyshyn, 1999). This is the case because of two reasons. Firstly, lexical influence on perception can adopt many seemingly conflicting forms, depending on the choice and timing of stimuli, the task, or demands on the working memory, that might conflate any linguistically driven top-down effect. Secondly, while there are many behavioural studies showing both facilitating and interfering effects of language on perceptual judgement, studies looking into neural correlates of those behavioural effects are still very few. However, with more research providing a nuanced profile of linguistic top-down influence on a wide range of perceptual tasks, using both behavioural and neuroimaging methods, a clearer picture is emerging about the nature of such effects.

This review aims to provide a summary of findings on the topic of linguistic influence on visual perception as reported in the literature, in light of neural mechanism underlying such processes. In order to do so, we will bring together literature from the field of grounded cognition and from research on linguistic (cueing) top-down influences on perception. We believe that both of these approaches complement each other in constructing a comprehensive narrative on how the linguistic and the perceptual systems interact. We therefore address the contribution from both line of research by putting forward the notion that any observed shift in how we receive and interpret visual input as a function of language is a by-product of linguistic grounding. Furthermore, we will relate these findings to more generally observed mechanisms underlying non-linguistic top-down (expectation and attention) influences in perception.

To that end, we will review studies looking at grounded cognition, using linguistic stimuli without a visual task or target, as well as studies using linguistic cues to modify the processing of visual input. We will specifically focus on studies of visual perception, rather than sensory-motor cortices (for the effects of language on sensory-motor systems, see Hauk & Tschentscher, 2013). Seeing as we are interested in the neural mechanisms of linguistically mediated perceptual activation, we will only look at studies that focus on the immediate, short-term, in the moment, effect of language on perception, rather than potential long-term effects of language (as in e.g., crosslinguistic studies, comparing speakers of languages with different semantic categories, see Wolff & Holmes, 2011).

2.2. Grounded Semantics

Researchers interested in the grounding of language in perception have mainly focused on studying the ability of language to spontaneously engage perceptual processes in the absence of visual inputs. Several studies have examined the neural correlates of semantic processing in a more naturalistic setting, such as passive listening to or reading stories, sentences or words (Huth et al., 2016; Mathôt et al., 2017, 2019; Nijhof & Willems, 2015; Saygin et al., 2010; Wallentin et al., 2011). They usually involved an eye-tracking or neuroimaging method, without a behavioural task, given that they do not set out to measure the effect of language on behaviour. The advantage of these studies is that the findings are not conflated with demands associated with the task. The disadvantage is that the researchers have little control over how much attention participants are paying to the stimuli during the experiment.

Using fMRI, it has been shown that listening to stories evokes neural activation patterns encompassing cortical surfaces far beyond conventional language comprehension and semantic integration areas, such as the left inferior frontal (IFG) or the middle temporal (MTG) gyri and the left anterior temporal lobe (ATL) (Huth et al., 2016; Nijhof & Willems, 2015). Not only did these studies show that linguistic inputs engage broadly distributed networks, they have also highlighted the fact that the neural representation of any given word is both modality-specific (i.e., different cluster can be observed for words depicting actions, tools, colours, etc.) and influenced by the company the words keep in any given context. However, there is still more to be examined about the neural mechanisms underlying linguistic ability to engage non-linguistic regions (for an overview of the shortcomings of these types of studies, see Barsalou, 2017). Additionally, some of these studies might have capitalised on the evocative power of rich linguistic contexts – stories – to recruit featural information from perceptual regions, raising the question of whether contextually poorer stimuli, like isolated sentences or words, could still engage visual cortices.

To that end, more controlled studies have used carefully constructed sentences in order to examine whether their comprehension engages perceptual processes (Saygin et al., 2010; Wallentin et al., 2008). Saygin et al. (2009) used fMRI to investigate whether motion sentences can modulate visual motion perception regions in the brain, such as the MT/V5 area, during natural language comprehension. They found a gradual shift in the activation of the MT/V5 region, with motion sentences exerting the strongest influence on the region, fictive motion sentences showing a slightly lesser effect and static sentences having no effect on the MT/V5 activation. These results show a nuanced profile of the ability of language to modulate visual regions, with even metaphorical concepts being grounded in the literal meaning of the words and their perceptually encoded features, albeit less strongly than literal sentences.

Looking at the effect of isolated words on the visual system in the absence of visual stimuli, studies have shown that even linguistic labels in isolation can engage perceptual cortices, such that the features encoded in the semantics of the labels activate regions close to or overlapping with the cortical areas that mediate the perception of those features (Martin et al., 1995; Rueschemeyer et al., 2010; Simmons et al., 2007). Mathot et al. (2017: original study in French, 2019: replication study in Dutch) measured participants' pupillary dilation and contraction in response to auditorily presented words conveying the notion of darkness (e.g., night, shadow) or light (e.g., day, sun). The authors observed pupillary dilation in response to darkness conveying words, and constriction in response to light conveying words, suggesting that a word's semantics modulates the activation of cortical areas responsible for pupil control in an involuntary manner (Mathôt et al., 2017, 2019).

Interpreting Results on Grounded Semantics

Taken together, these studies show us that when reading or listening to linguistic stimuli, language spontaneously engages functional networks that surpass conventional language regions in temporal and frontal cortices. Specifically, they have demonstrated that linguistic stimuli, from elaborate stories to words in isolation, can recruit perceptual cortices and activate regions that usually respond to visual stimuli, but in this case, they respond to the visual representations of the features encoded in the word semantics.

While these studies show that language can modulate neural activation in perceptual regions, even in the absence of a visual stimulus, they cannot tell us what the behavioural and neural consequences of such modulations are in the presence of a visual input. In other words, they cannot tell us how these modulations change our perceptual experiences and the way in which we encode incoming visual input. In order to tackle that question, we next look at the studies examining whether and to what extent language can modify how we receive and interpret visual input and how those modifications are actualised in terms of neural activation patterns.

2.3. Cueing Effects of Language on Perception

Studies interested in the ability of language to modify ongoing perception and perceptual judgements usually utilize a cueing paradigm. In this paradigm, people are presented with a linguistic stimulus, such as a word, sentence, or short story, prior to doing a perceptual task, such as detection, discrimination, or a visual search task (i.e., identifying the target among multiple distractors) on the target

(Boutonnet & Lupyan, 2015; Dils & Boroditsky, 2010a; Francken , Kok, et al., 2015; Meteyard et al., 2007; Slivac et al., 2021). By intercepting the process of language comprehension with a visual input, these studies can directly examine how semantic grounding, observed in studies with purely linguistic inputs covered above, affect the processing of visual information, both behaviourally and neurally. The benefit of these studies is that they, by virtue of having an explicit task, have more control over what participants pay attention to. Additionally, by connecting behavioural effects with the neural pattern of activation in response to experimental conditions, they provide a more comprehensive account on language-perception interaction and tap into possible mechanisms that lead to those interactions. The disadvantage of such experiments is that the task may conflate the effects of the conditional manipulation on behaviour (Hoenig et al., 2008; Kerzel et al., 2009; Yee et al., 2012; Yee & Thompson-Schill, 2016).

2.3.1. Behavioural Studies

Studies investigating how language comprehension affects our visual processing have reported both facilitating effects on target perception, when language cues are congruent with the target, and detrimental effects in the case of cue-target incongruence, suggesting that language has the power to change our perceptual experiences, i.e., modify how we receive and interpret visual input.

Studying the effect of motion stories on visual motion perception**,** Dils and Boroditsky (2010a) found that listening to linguistic descriptions of motion induced mental images strong enough to cause a motion aftereffect illusion – an illusion of a directional motion in an static or incoherent RDM stimulus caused by prior exposure to motion in the opposite direction (Dils & Boroditsky, 2010a). This finding demonstrated that linguistic processing can change how we perceive subsequent visual input, to such an extent that they can induce an illusion usually only triggered by visual stimuli. These studies again profited from rich linguistic input, suggesting that the potential visual shift could have been the product of participants' imagery in response to language abundantly reinforcing a categorical feature (e.g., upward motion).

Studies using more concise linguistic stimuli, have also investigated the effect a single sentence can have on perception of visual stimuli. These studies observed faster and/or more accurate responses to visual stimuli matching the content of the previously read sentence, such as motion direction, object orientation and shape (Pelekanos & Moutoussis, 2011; Stanfield & Zwaan, 2001; Zwaan et al., 2002).

Using still more concise lexical stimuli – single words, i.e., linguistic labels – studies have observed some diverging results. Using linguistic labels as spatial cues, some studies have confirmed the label advantage in the form of faster reaction times for spatially congruent target detection (Dudschig et al., 2012) and discrimination (Ostarek & Vigliocco, 2017). Conversely, other studies have shown that language spatially congruent with the target can have an inhibitory effect in the form of slower responses to the target (Estes et al., 2008; Richardson et al., 2003; Verges & Duffy, 2009). Such effects have been explained by arguing that the process of identifying a target in the cue-congruent location may require an inhibition of spatial features activated by the cue first (Estes et al., 2015). It has also been hypothesized that the inhibition effect in those studies is due to the feature overlap being manipulated between the cue and the location of the target, rather than the cue and the target itself (Dunn et al., 2014), or due to the fact that the cueing effect interfered with the more demanding discrimination rather than a simpler detection task (Dudschig et al., 2012). However, given that the two studies that reported the cue-induced spatial facilitation effect used spatially congruent but cue-irrelevant targets (Dudschig et al., 2012) and a discrimination task (Ostarek & Vigliocco, 2017), more research is needed in order to understand the conditions and mechanism behind the inhibition effects of cues spatially congruent with the target location.

Moving beyond spatial cues, numerous studies have demonstrated that the presentation of a single word (i.e., lexical cue) shortly before a visual target can improve (in the case of cue-target congruence) or interfere with (in the case of cue-target incongruence) target discrimination (Forder & Lupyan, 2019; Lupyan & Thompson-Schill, 2012; Ostarek & Vigliocco, 2017), detection (Bidet-Ildei et al., 2011; Edmiston & Lupyan, 2015; Francken, Kok, et al., 2015; Meteyard et al., 2007; Pavan et al., 2013), or visual search (Lupyan & Spivey, 2010b; for review see Huettig et al., 2011).

This effect has been shown to persist even when the visual target is rendered imperceptible or ambiguous by means of masking techniques such as continuous flash suppression (CFS; Forder et al., 2016; Lupyan & Ward, 2013; Ostarek & Huettig, 2017; Pinto et al., 2015; Stein & Peelen, 2015), random dot motion masking (Bidet-Ildei et al., 2011; Slivac et al., 2021) or thresholding of motion coherence levels (Francken, Kok, et al., 2015; Meteyard et al., 2007; Pavan et al., 2013). The design of these studies allows researchers to examine the extent to which we depend on higher level systems to fill in the gaps in our perception, and the results emphasize the reliance on the top-down linguistic information for the successful target perception in visually noisy environments.

Interestingly, few studies have found the linguistic effect on perception to be left-lateralised, i.e., observed when stimuli were presented to the right visual field, but not left (Francken, Kok, et al., 2015; Zhou et al., 2010). Given that language processing itself is left-lateralised, these findings suggest that strength of the linguistic effects on perception may be dependent on the (hemispheric) proximity of the perceptual region to the linguistic areas. So far, these results are in line with studies reporting that long-term effects of categorisation as learned through language on perception are also left-lateralised (Drivonikou et al., 2007; Gilbert et al., 2006; Mo et al., 2011; Regier & Kay, 2009).

Further, studies have shown that even cues unrelated to the task (Lupyan & Thompson-Schill, 2012; Slivac et al., 2021) or rendered unaware by means of masking (Francken et al., 2015) can modify how we process visual input. Additionally, Slivac et al. (2021) have shown that even cues unrelated to the task or the target, but rather congruent with the distractor stimulus, can still be detrimental to the detection of the target, suggesting that the distractor cues were strong enough to induce a bias away from the target. However, the cueing effects observed with subconscious or task irrelevant cues are less pronounced than those with task-related or overtly shown cues, suggesting that the clear cue-task relatedness can reinforce the extraction of the features encoded in the cue semantics.

However, opposite claims have also been made. Studies have pointed out that contextual and task demands can substantially change the type of information extracted from the word meaning (Hoenig et al., 2008; Yee et al., 2012; Yee & Thompson-Schill, 2016). Taken together, the evidence suggests that a task closely connected to the cue semantics can alleviate computational burden by prioritizing processing of that subset of lexical information deemed to be of the highest relevance for the task. While lexical influence on perception can survive task deviations from the semantic information encoded in the cue, it is still unclear to what extent shifting attention away from the target brings about a more substantial change in the kind of features that end up being extracted from the linguistic labels.

In sum, current behavioural evidence shows that the linguistic grounding can shift how we receive and interpret visual inputs. Namely, hearing or reading linguistic cues featurally congruent with a visual target facilitates the perception of that target, and it can even boost the target detectability threshold (i.e., make the invisible targets visible). Conversely, in the case of cue-target incongruence, i.e., lack of shared features, language has a detrimental effect on our ability to process visual information quickly and accurately. Current evidence also shows that the choice of task can up- or downregulate the potency of lexical effects on perception by highlighting the relevance of the semantic content conveyed by the cue needed for the successful performance on the task.

Interpreting Behavioural Results

Linguistic influence on perception can manifest itself in a number of different ways. Cueing studies using a visual detection or discrimination task investigating this phenomenon have interpreted findings pertinent to accuracy scores, reaction times, and signal detection theory indices (discriminability/sensitivity and bias) as proof of linguistically mediated facilitation of perceptual processing.

To illustrate, comparing cue-target congruent with incongruent experimental conditions, motion perception cueing studies found a facilitating effect of congruent motion cues on motion perception, as reflected in both an increase in accuracy and faster reaction times (Francken, Kok, et al., 2015; Slivac et al., 2021), faster or slower reaction times only (Bidet-Ildei et al., 2011; Richardson et al., 2003), a shift in both d’ and Criterion (Meteyard et al., 2007), and a shift in Criterion only (Francken, Kok, et al., 2015; Slivac et al., 2021). Although these findings may seem somewhat inconsistent, the explanation for this inconsistency may lie in the experimental settings themselves.

One possible explanation for some of these discrepancies in findings come from the study done by Pavan et al. (2013), who tested if and how lexical effect on perception changes as a function of an increase in target ambiguity. Presenting participants with random dot motion targets with coherence levels at suprathreshold (84%) and threshold (50%) accuracy levels, they found a double dissociation between discriminability (d’) and RTs for the two types of targets. For the suprathreshold condition, the RTs were faster for directional cue-target congruence, but there was no difference in discriminability. For threshold condition, no RT effects were found for cue-target congruence, but there was a shift in d’ (Pavan et al., 2013).

The theoretical explanation for this dual dissociation may be that highly detectable stimuli do not necessitate reliance on the top-down influence for a successful target detection, i.e., priors have the strongest influence on perception when visual input is ambiguous (Bogacz et al., 2006). Additionally, given the high accuracy performance on tasks with highly visible targets, there may not be much space to significantly increase accuracy as a function of facilitating cues compared to the baseline condition, especially given attention lapses and time constraints regularly present in cueing experiments. In those cases, the facilitation might manifest itself in reaction times instead. However, in the case of ambiguous stimuli, our reliance on top-down information for target perception becomes much more meaningful. In addition, there is a lot more space to improve performance on the task as a function of congruent cues. However, the time necessary to resolve ambiguity might not be significantly changeable within the restricted time window that cueing detection studies usually have. More comprehensive measures such as speed-accuracy trade-off can potentially provide further insight into how the brain processes visual information as a function of congruent and incongruent lexical cues.

Additionally, signal detection theory indices, such as sensitivity/discriminability (most often expressed as d') and bias (most often expressed as Criterion) measures are often reported as being indicative of the cognitive and neural loci of the linguistic top-down influence on perception. Sensitivity measures tend to be interpreted as a sign of perceptual modulations, while bias has been taken to be an indicator of higher-level, decision-making processes (Meteyard et al., 2007; Pelekanos & Moutoussis, 2011). This dichotomy, however, has been brought into question by a number of studies, which pointed out that signal detection indices must be interpreted in the context of experimental settings, such as stimuli and task (Georgeson, 2012; Witt et al., 2015). Furthermore, it has been shown that shifts in bias are accompanied by modification in activation patterns in perceptual regions. Several studies have presented convincing results showing that Criterion in discrimination tasks is either sensory in nature, or both sensory and decisional, but not solely decisional (Linares et al., 2019; Morgan et al., 1990; Ratcliff et al., 1989).

Even knowing that both sensitivity and bias shifts stem from activity in visual cortices, the discriminability-bias dichotomy does have an important role – it can tell us much about the behavioural consequences of lexically induced perceptual shifts. One possibility, reflected in changes in discriminability, would be that linguistic activation of visual regions helps us differentiate between two stimuli, i.e., make us more adept at distinguishing between the target and non-target. Another possibility, reflected in bias shifts, is that language warps visual perception towards the cued concept, such that under the influence of a target-congrunet cue, we are less successful at discriminating between two stimuli, and more likely to classify a non-target as target, especially in ambiguous settings.

Summerfield et al. (2016) proposed that the distinction between discriminability and bias can be connected to the feature-based attentional and expectational top-down influences. They claim that feature-based attentional top-down processes manipulate the relevance of the incoming visual signal, i.e., they upweight the information relevant for the task during the decision making process, resulting in discriminability (i.e., sensitivity) shifts. Feature-based expectation, on the other hand, affect the probability of the incoming signal, thus adjusting the decision criterion towards the more probable features, i.e., those that occur more frequently in any given context (Summerfield & Egner, 2016). Even in non-linguistic studies on top-down influences on perception, these two aspects have rarely been explicitly orthogonalized. In studies involving language, this is particularly difficult to achieve, given that language semantics is a complex and multifaceted medium towards conceptualisation and feature extraction.

2.3.2. Neuroimaging and Eye-Tracking Studies

While studies on grounded cognition have relied mostly on neuroimaging techniques in their experiments, cueing studies with perceptual tasks, however, have mostly employed behavioural paradigms, with very few attempts to study neural activation patterns accompanying such perceptual shifts. Still, the presence of eye-tracking and neuroimaging techniques is not completely absent, so we will take a look at the current state of knowledge about the neural correlates of linguistic effects on perception.

These studies tackle the question of whether linguistic top-down processes change the neural representation of visual stimuli in perceptual areas, i.e., whether the integration of lexical information and bottom-up visual inputs is already observable at earlier levels of perceptual processing. Alternatively, early perceptual processes may be unaffected by top-down lexical influences and the integration of lexical information and perceptual processing may be happening at the higher, decision-making regions instead.

Eye-tracking studies using words conveying spatial information, have shown that linguistic cues, implicitly encoding the spatial feature congruent with the target location, can facilitate saccadic eye movements (initiation and speed) towards that location (Dudschig et al., 2012, 2013; Dunn et al., 2014). These results suggest that words may reach neural processes responsible for the control of eye movements and pupillary responses.

EEG measurements, given high temporal resolution, can tell us whether the label advantage stems from modulations in early visual processing level or later, semantic and decision-making level, by looking at whether lexical cues elicit early visual ERP components, such as the P100 and N100, or a later ERP component, such as the N400, associated with higher-level semantic integration (Boutonnet & Lupyan, 2015; Landau et al., 2010; Noorman et al., 2018). Landau et al. (2010) examined the temporal dynamics of the linguistic effect (sentences) on face perception and found that the magnitude of the N170, associated with face perception, was larger after face-describing sentences than after scene-describing sentences, and only for the left hemisphere. Looking at the top-down effects of linguistic labels on perception, Boutonnet & Lupyan (2015) found that hearing a word affected early visual processes. Namely, they found that the P100 was larger when participants were cued by labels (e.g., the word *dog*) compared to equally informative nonverbal cues (e.g., dog bark), with the enhancement occurring within 100ms of image onset. Similarly, Noorman et al. (2018) showed that hearing a word activates representations of its referent's shape, which interacts with the visual processing of a subsequent picture within 100 ms from its onset (Noorman et al., 2018). These findings are compatible with the general view that features extracted from cues act as immediate priors for the visual system, thus immediately biasing the reception and processing of the incoming visual target.

Using sentences with content matching or mismatching perceptual features of target objects in an MEG study, Hirschfeld et al. (2011) found two effects: an early modulation of the occipital cortex within 120ms, reflecting early visual processing, and a later modulation in the N400 window in the left temporal cortex, sensitive to higher level processes such as lexical access and semantic integration. These results show that the loci of the linguistic top-down influence on perception can be detected at both lower perceptual levels, as well as higher semantic levels.

Studies using fMRI have looked at the extent to which lexical cues can modify neural activation in perceptual regions responding to the visual target, as reflected in changes in the BOLD signal (Francken, Kok, et al., 2015; Pirog Revill et al., 2008; Puri et al., 2009). Puri et al. (2009) employed a cueing task in an fMRI setting to investigate whether linguistic labels (face vs. place) cue the perception of faces and places, known to be represented in the fusiform face area (FFA) or parahippocampal place area (PPA), respectively. They found that cue-target congruence led to an increase in the activity in the target encoding regions, compared to the incongruent pairings. These findings showed that linguistic labels can modify neural activity in cortical areas responsible for encoding visual features, which in turn suggests that language can cascade into the perceptual system and modify neural activity therein. Conversely, Eger et al., (2007) used degraded visual stimuli to examined the effects of lexical cues on object perception, and found that even though objects congruent with the cue were recognized earlier than incongruent ones, when controlling for recognition point and stimulus information, activity in the ventral visual cortex reflected recognition success, independent of the cueing condition. Similarly, Francken, Kok, et al. (2015) looked at the cueing effects of motion words on the perception of motion directionality, and found the locus of the linguistic effect on motion perception in the middle temporal gyrus (MTG), involved in semantic integration, rather than the hypothesized MT/V5 region involved in motion perception. Taken together, fMRI research presents a divided picture, without clear evidence for whether linguistic influence on perception can be detected in earlier visual cortices using the fMRI technique.

Interpreting Neuroimaging Results

Neuroimaging cueing studies show divergent results between M/EEG and fMRI results. Namely, M/EEG studies are consistent in showing early influence of linguistic cues on perception, while this observation is largely absent from the fMRI studies. This is particularly surprising given that studies on grounded semantics reliably show an engagement of perceptual regions corresponding to visual representations of categories and features depicted by language.

While there is a possibility that any visual input overpowers the grounded semantic neural effects, M/EEG and behavioural findings suggest that this is not a likely scenario. We should note that none of the fMRI cueing studies reported above had a control condition without a lexical cue (i.e., all clues, congruent, incongruent and neutral, when present, are lexical). One possible scenario, therefore, is that lexical cues do shift neural activation in perceptual regions compared to non-lexical cues, but the difference in cue congruence between different lexical cues themselves is not reflected in an overall increase or decrease of activation within a visual region of interest, as measured with standard univariate fMRI analyses approaches. Further, the effect of lexical cues congruent with the target, compared to those incongruent with the target, could only be detectable by looking at the differences in more fine-grained activation patterns within perceptual regions of interest, detectable with multivariate pattern analyses approaches (i.e., looking at subtle shifts in neural responses that take into consideration the patterns of activation across voxels).

fMRI studies investigating top-down influences of non-linguistic cues on perception have shown that top-down cueing effects lead to suppressed responses to the bottom-up signal incongruent with the cues while simultaneously enhancing the response to expected, i.e., cue-congruent, features. This interplay between dampening of unexpected and enhancing of the expected stimuli results in the so called *sharpening* pattern of activation within a region in response to cued features, rather than an increase in the average magnitude of activation (Kok et al., 2012; Kok & de Lange, 2014; Lee & Mumford, 2003). In Bayesian terms, the attenuation, and by extension sharpening, accounts have been explained by the findings that the subpopulation of neurons coding for prediction error signals diminish, while those coding for expected stimuli (predictions) enhance, as a function of the expectation build-up, leading to an overall sharper representation of expected stimuli. The sharpening activation profile is not observable with the standard, univariate fMRI analysis, which looks at the change in the average magnitude of activation within a region by evaluating each voxel in isolation, rather than patterns of activation within a region that take into consideration joint activation of multiple voxels. While the sharpening account of top-down influences on perception has been reported and argued for in a number of studies using non-linguistic stimuli (Kok et al., 2012; Kok & de Lange, 2014; Martens & Gruber, 2012; Yon et al., 2018; K. Zhang & Sejnowski, 1999), it has yet to be tested in studies with linguistic labels as cues.

To summarise, M/EEG studies looking into language-perception interaction reliably show that linguistic cues can affect low-level visual systems as soon as 100ms post target presentation. fMRI studies, however, are more divided when it comes to neural activation correlates of language-perception interaction. Not many studies have been done on the topic, however, and those that have been done report univariate results only, which can only measure the increase (or decrease) in the average magnitude of the signal in any given region, but tell us nothing about the more subtle shifts in activation within those regions. Non-linguistic cueing studies, on the other hand, have shown that the effects of top-down expectations on perception may be accompanied by the sharpening rather than uniform increase or decrease of activation in early visual regions. This suggests that the fMRI studies may have been searching for the wrong neural signature of the behaviourally observed linguistic cue-target interaction effect.

2.4. Discussion

In this review, we have provided an overview of the current state of knowledge regarding linguistic effects on perception, by summarising findings from the literature investigating grounded cognition and linguistic (cueing) top-down influences on perception. In the introduction, we have proposed a claim that linguistic top-down influences on perception observed in cueing studies are a by-product of semantic grounding, i.e., that observations from the cueing studies look at how the process of grounding modify how we receive and interpret visual input.

The grounding literature has reliably shown that the processing of linguistic stimuli, from stories and sentences to individual words, can activate perceptual regions responsive to visual representations of categorical features conveyed by lexical items. These findings are in line with the mechanisms proposed from both linguistic and non-linguistic top-down influences on perception: perceptual system can be activated by cognitive (in this case linguistic) processes in a feature-based manner (Lupyan, 2012; Summerfield & Egner, 2016). In other words, part of the visual system responsible for the encoding of visual features representative of a category can also be engaged in the absence of visual inputs, during semantic processing of that category. But these findings beg the following question: how does spontaneous engagement of perceptual system by means of grounding affect how we process visual information?

Behavioural cueing studies with visual targets have shown that this process of feature-based semantic grounding can change how we receive and interpret visual inputs. The exact behavioural consequences of that influence, however, are not as consistent as one would think just looking at the literature on grounded semantics. In particular, studies investigating the effect of cue-target spatial congruence on target detection and discrimination have reported both facilitatory and inhibitory reaction times as a function of spatially congruent cues (Estes et al., 2008; Ostarek & Vigliocco, 2017; Richardson et al., 2003). Various studies have tried to provide an explanation for the inhibitory results (see above), but the crucial reason for this discrepancy seems to be the extent to which the cue and target overlap in categorical features.

These findings illustrate what we propose in this review: in studies reporting inhibitory effects, the grounding happening as a function of cues was in the location, not in the target itself, meaning that these studies overlapped two visual processes: the spatial grounding put forward by the cue and visually reinforced by the target location, and the processing of the visual target itself, unrelated to the cue or the features characterising the location. In this situation, the spatial encoding evoked by the cue would indeed need to be inhibited before the encoding of the target (i.e., an unrelated visual input) can happen, causing a delay in the processing of that target.

Studies using cues featurally overlapping with targets, on the other hand, present a consistent body of evidence showing that lexical cues can facilitate the perception of the visual target. However, the question remains about the neural signature of such facilitation. One possibility is that linguistic cues can modify the activation in early visual regions. Another possibility is that the behavioural facilitatory effects represent higher-level cognitive and decision processes. Behavioural cueing studies have employed signal detection theory approaches to make claims about the locus of the lexical cueing effect, associating the sensitivity measure with lower-level perceptual modifications and the bias measure with higher-level, decision processes. We argue that this approach is misleading and has not been supported by studies directly looking at the neural corelates of biases, as defined by signal detection theory (Linares et al., 2019; Morgan et al., 1990; Ratcliff et al., 1989). Namely, while we recognise the value in distinguishing shifts in bias from shifts in sensitivity in terms of behavioural consequences, we argue that both can be accompanied by shifts in neural activation in perceptual regions, and are therefore not well suited for making claims about the locus of the effects.

Adopting a more direct approach to examining the neural loci of activation, studies on grounded semantics are fairly consistent in their findings, whereas cueing studies with visual targets are divided on the topic. M/EEG cueing studies have reported the modulation of processes in the early visual cortex as soon as 100ms post target presentation (Boutonnet & Lupyan, 2015; Hirschfeld et al., 2011; Landau et al., 2010; Noorman et al., 2018). These results follow the reports from the grounded cognition fMRI literature, in showing that linguistic effect on perception can be observed at the early stages of visual processing. They are also in agreement with studies investigating more general, non-linguistic feature-based top down influences on perception (Dunovan et al., 2014; Gong & Liu, 2019; Liu et al., 2007; Maunsell & Treue, 2006; Saenz et al., 2002; Summerfield & Egner, 2016).

However, evidence coming from fMRI cueing studies stand in contrasts to both fMRI studies on grounded semantics and M/EEG cueing studies. One reason for that could be that these studies have been looking for the wrong or incomplete signature of the effect. The few studies done on the topic so far have reported only univariate fMRI analyses, whereas we propose that multivariate approaches might be more appropriate to shed light on the more subtle neural signatures of language-perception interaction. Specifically, we propose that linguistic cueing studies need to directly test the existence of the sharpening account of neural activation, proposed by non-linguistic cueing studies (Kok et al., 2012; Kok & de Lange, 2014).

Furthermore, a clearer picture is needed on the condition that would lead to inhibitory cueing effects. Current accounts explain the inhibition by claiming a lack of feature-based overlap between cue and target, however, some studies also bring up the issue of task difficulty. The relationship between the (lack of) feature overlap and inhibition or enhancement effects would therefore need to be directly tested in an experimental setting, while keeping the task constant.

Finally, the neural correlates and experimental settings leading towards bias or sensitivity shifts should be directly measured. This would help us understand under which circumstances top-down lexical influence biases our perception towards what is labelled, and under which circumstances it aids in our ability to discriminate between labelled and non-labelled percepts.

Further research addressing these remaining issues can delineate which cognitive and neural mechanisms support linguistic top-down influences on perception, and help connect those mechanisms to more general processes underlying top-down expectation and attention effects on perception.

CHAPTER 3

Linguistic Labels Cue Biological Motion Perception and Misperception

Abstract

Linguistic labels exert a particularly strong top-down influence on perception. The potency of this influence has been ascribed to their ability to evoke category-diagnostic features of concepts. In doing this, they facilitate the formation of a perceptual template concordant with those features, effectively biasing perceptual activation towards the labelled category. In this study, we employ a cueing paradigm with moving, point-light stimuli across three experiments, in order to examine how the number of biological motion features (form and kinematics) encoded in lexical cues modulates the efficacy of lexical top-down influence on perception. We find that the magnitude of lexical influence on biological motion perception rises as a function of the number of biological motion-relevant features carried by both cue and target. When lexical cues encode multiple biological motion features, this influence is robust enough to mislead participants into reporting erroneous percepts, even when a masking level yielding high performance is used.

Published as: Slivac, K., Hervais-Adelman, A., Hagoort, P., & Flecken, M. (2021). Linguistic labels cue biological motion perception and misperception. Scientific Reports, 11: 17239. doi:10.1038/s41598-021-96649-1.

3.1. Introduction

Perceptual systems are susceptible to a wide variety of top-down influences (Gilbert & Li, 2013; Summerfield & de Lange, 2014; Vetter & Newen, 2014), among which linguistic labels have been found to be particularly powerful. Numerous studies have demonstrated that the presentation of a lexical cue shortly before a visual target can improve discrimination (Forder & Lupyan, 2019; Lupyan & Thompson-Schill, 2012) or detection (Bidet-Ildei et al., 2011; Edmiston & Lupyan, 2015; Francken, Kok, et al., 2015; Meteyard et al., 2007) of that target. This effect has been shown to persist even when the visual target is rendered imperceptible via masking techniques such as continuous flash suppression, suggesting that any observed perceptual activation in these experiments is top-down in nature and in many cases label-driven (Forder et al., 2016; Lupyan & Ward, 2013; Ostarek & Huettig, 2017).

Lexically mediated cueing appears to be uniquely effective in eliciting perceptual activation changes, compared to similarly informative non-linguistic cues, such as environmental sounds (Lupyan & Thompson-Schill, 2012), suggesting that linguistic labels have privileged access to a substrate relevant to perception. The *label feedback hypothesis* formalised this supposition, positing that while non-linguistic cues, such as sounds or pictures, are necessarily exemplar-bound, linguistic labels activate categorical representations by abstracting from the idiosyncrasies of individual category members and emphasising the *diagnostic features* of that category (Lupyan, 2012). In doing so, they are able to activate a perceptual template that effectively warps the neural activation towards the labelled category (Boutonnet & Lupyan, 2015; Lupyan et al., 2020). This lexically induced perceptual activation towards the diagnostic features of the labelled category has been argued to occur in an automatic, task-independent manner, as exemplified by the fact that cueing effects on perception are reported across different types of tasks involving perceptual limens, and even when labels are task-irrelevant (Mathôt et al., 2017). However, Klemfuss et al. (2012) caution against the claims that language can modulate perceptual activation, and propose an account according to which the linguistic cueing effect reflects reduced burden on working memory (Klemfuss et al., 2012).

However, visual perception extends beyond the commonly examined detection of static objects, which relies on contour (i.e., shape or form) recognition (for a discussion on stimulus and task complexity in cueing paradigms, see Kerzel et al., 2009). It also includes motion perception. Motion is inherently dynamic and transient, and perceiving it requires integration of relevant elements over space and time for a successful construal of a coherent percept. A small number of studies have used dynamic, point-light stimuli, such as random dot motion (RDM), to examine the influence of lexical cues on the perception of motion. Identification of RDMs' direction of motion requires integration of the constituent elements (position and kinematics of individual dots), into a coherent, directional percept. Using RDMs set at the coherence decision limen, it has been shown that motion verbs (e.g., *rise*, *fall*) as linguistic cues bias the perception of RDMs – verbs congruent with direction of the dominant motion vector of RDMs facilitate the judgement of its principal motion direction, while incongruent verbs reduce the accuracy of such judgement (Francken, Kok, et al., 2015; Meteyard et al., 2007). These studies provide evidence that the perception of dynamic stimuli is susceptible to linguistic influence.

Such results raise the possibility that even more complex, dynamic point-light stimuli are subject to lexical cueing effects. One such class of stimuli is point-light figures (PLFs), used to study biological motion perception (Johansson, 1973). The perception of PLFs is *compositional*, in that it requires the observer to compose disconnected dots representing bodily joints and their local kinematics into a unified percept of a human figure in action ('form-from-motion' stimuli; Boxtel & Lu, 2015). The category of biological motion encompasses a wide variety of different actions performed by biological entities (e.g., for humans, *walking* or *cycling*) the perception of which requires an integration of several different features, most diagnostic (or defining) of a particular type of action being form and kinematics characteristic of that action. In the case of PLFs, these features point to different aspects of the stimulus assembly into the global percept: local information on kinematics is given by the individual dots in the earlier stages of target composition, while global form information emerges upon the successful binding of the dots into a recognisable figure (George Mather et al., 1992; Thirkettle et al., 2009; Thornton et al., 1998). Presenting such stimuli in combination with lexical cues conveying the notion of form and kinematics, individually or in combination, can reveal how different aspects in the process of target configuration are affected by cues encoding different features relevant to biological motion.

For example, in relation to the concept of biological motion, the word *brother* makes the feature of human form (i.e., human body) directly available to us, but carries no information about an action the named entity might be engaged in. The word *rower*, on the other hand, makes both the information about the human form (i.e., human body) and the particular type of kinematics necessary to perform such action (sitting position with a characteristic arm movement) directly available to us.

This feature-based conceptualisation of the content of linguistic labels in combination with form-from-motion PLF stimuli allows us to test the hypothesis that the efficacy of lexical top-down influence on perception lies in the ability of those labels to highlight and activate conceptual and perceptual representations of category-diagnostic features, and bias the perception of the visual input towards the labelled category. In this study, we therefore examine how the encoding of biological motion features – form and kinematics – in linguistic labels, modulates the strength of linguistic influence on the perception of biological motion represented in PLFs performing an action. In order to achieve this, we manipulate the number of biological features encoded in linguistic labels (no feature: no biological form or kinematics, single feature: biological form only, or multiple features: biological form and kinematics), and their degree of overlap with the PLF targets performing an action (Experiment 1), as well as congruence between label and target for cues with multiple feature availability (Experiment 2) in a biological motion detection task. To reject the possibility that any cueing effect depends solely upon visual form detection (only biological form feature encoded in the visual target), we also test whether lexical cues can affect the orientation discrimination of not only naturally moving PLFs, but also those captured in a recognizable, action-characteristic frame and stripped of the local kinematics feature, but moving horizontally in a rigid manner (‘gliding’, figures; Experiment 3).

Our view of the mechanism underlying the supposition made by the label-feedback hypothesis in the context of the current study is the following: form and (when encoded) kinematics features, when delivered lexically, co-activate perceptual form and kinematics representations relevant to the named action. This activation is not an all-or nothing phenomenon, but rather gradually becomes more extensive with the number of target relevant features encoded in the label. Cues impoverished with respect to the visual target, with only one (form-only) feature encoded, would therefore fail to evoke a comprehensive perceptual template necessary for target recognition (form *and* kinematics encoding neurons), and as such may exert only a weak influence on target perception. In other words, because the form-only, lexically induced bias is not strong, congruence will not give the perceptual system awaiting visual target a strong initial boost, but it will also allow it to ‘recover’ more quickly in the case of cue-target mismatch. In the case where lexical cues encode more features (both form and kinematics), they will engage a more comprehensive conceptual and perceptual representation, reflecting both form and kinematics encoding. By doing so they will bias the perceptual activation more strongly towards the labelled category and as a result exert a stronger influence on the perception of the incoming target in the following way: When congruent, the (pre-)activated neurons overlap with those that would need to be activated for target perception, such that they are already ‘firing’ by the time the visual input arrives (the activation is already ongoing). When incongruent, they will derail target perception, because the induced pattern of activation does not overlap with that necessary for target perception, i.e., the ongoing lexically induced activation is thus misleading with respect to the target, and needs to be corrected for successful target perception (uninformative or ‘mismatched’ template needs to be supressed while the target matching one needs to get activated). In this featurally more comprehensive case (i.e., cues with multiple features), therefore, the congruent lexical boost will be stronger compared to that exerted by featurally impoverished (single feature) cues, but the recovery period in the case of the cue-target mismatch will also take longer or be harder to achieve withing a short time period.

If feature activation drives lexical cueing effect on perception, the immediate availability of both form and kinematics features carried by lexical cues is hypothesized to exert a stronger influence on biological motion perception, compared to cues with single (form) feature availability: when congruent with the target, we expect them to amplify visual detection and interfere with rejection; when incongruent with the target, we expect them to interfere with visual detection and facilitate rejection.

In other words, we expect that the congruent biological motion cues will lead to an overall, conceptual and perceptual, bias towards the labelled category, which will result in a shift in Criterion: higher detection (hit) rate on trials with coherent PLFs, but also higher false alarm rate on trials with scrambled PLFs, with participants wrongly thinking they are seeing what has been prompted by the cue. This bias is hypothesized because even in the absence of the human form, we expect the kinematics feature preserved in the scrambled PLFs to overlap enough with the representation prompted by the cue to mislead participants into wrongly composing the scrambled PLF dots into a coherent percept. In the case of incongruence, the overlap with the target is absent when it comes to the kinematics feature, which is a particularly important clue for action recognition among the masking dots, so we expect a decrease in hit rate and an increase in correct rejection rates.

Given our experimental design and hypotheses, i.e., we expect both the target coherent distribution and the target scrambled distribution to shift as a function of our cues, we will interpret any such shifts in our results as reflecting conceptual and perceptual bias (cf.(Witt et al., 2015)). In other words, we do not ascribe the bias induced by the cues to one single processing level. While we do think that

Criterion scores can indicate the participant's strategy (decision or response level), we echo previous accounts stating that this is not the only bias that Criterion reflects. As has been argued before (Balakrishnan, 1999; Georgeson, 2012; Linares et al., 2019; Witt et al., 2015), we recognise the necessity to interpret SDT indices in line with the experimental design when conducting psychological detection or discrimination experiments. The abundant evidence showing that linguistic top-down influences can and do regularly modify conceptual and perceptual alongside higher level decision processes (Boutonnet & Lupyan, 2015; Dils & Boroditsky, 2010a, 2010b), and that those modifications affect Criterion scores (Linares et al., 2019; Morgan et al., 1990) further justifies our claim that the bias observed here should not be restricted only to the decision making level, but also encompasses conceptual and perceptual levels.

Finding that a lexical cueing effect is modulated by the overlap in the amount of features encoded in lexical cues and visual targets would be the first empirical demonstration of a process of feature activation as underlying such linguistic influences on perception.

3.2. Methods and Results

3.2.1. Experiment 1

We examine whether lexical cues carrying and overlapping in multiple features diagnostic of biological motion category (form and kinematics) exert a stronger effect on the PLF detection, compared to single-feature overlap (form-only) conveying cues, as well as no-feature overlap (general motion cues, that had overlapping features with the mask rather than the target; see details below).

Participants

Fifty-one native speakers of Dutch (43 female, 8 male; mean age: 23.56; age range: 19-33) recruited from the Max Planck Institute (MPI) participant database took part in the experiment. Eleven participants failed to reach the inclusion criterion during the thresholding procedure (see below) and were therefore excluded from the analysis, resulting in 40 complete datasets (33 female, 7 male participants; mean age: 23.38; age range 19-33). All participants were right-handed and had normal or corrected-to-normal vision, and no reading difficulties. All the participants gave their informed consent and received financial compensation for their participation. All the studies presented in this article were approved by the Ethics Board of the Social Sciences Faculty of Radboud University (ESCW). All experiments were carried out in accordance with the recommendations of the seventh revision of the Declaration of Helsinki (2013) regarding participants' informed consent.

Stimuli

All stimuli were generated using the Psychophysics Toolbox(Brainard, 1997) within MATLAB R2016a (MathWorks, Natick, MA). Both lexical cues and visual targets were presented in white (luminance: 160 cd/m^2) on a grey background (luminance: 37 cd/m^2).

The lexical cues were presented in Dutch and consisted of three lexical cue categories with 4 nouns each, and one control (no language) cue category, the string #### (see **Table S1**). The three categories of lexical cues encompassed two categories semantically congruent with the target, and one semantically incongruent category. The former contained biological motion cues, conveying both biological form and kinematics information (e.g., *rower, walker*), and biological form cues, with biological form but not motion information (e.g., *brother, father*). Semantically incongruent cues were general motion words (e.g., *snow, smoke*), which matched the directionality of the masking RDM dots on every trial (e.g., snow – downward motion, smoke – upward motion) rather than the PLF target.

The visual targets consisted of 13 white dots comprising a point-light figure (PLF; size: 3.59 – 4.36cm (horizontal)*5.95 – 6.57cm (vertical); speed: 30 frames/second) embedded in a random dot motion mask (RDM), with circular aperture (22*23.5cm; number of dots: 866; dot size: 0.528*0.528mm, dot motion speed: 0.528mm/frame, dot lifetime: 10 frames, at 30 frames/second), presented in the middle of the screen. Four PLF types, performing four types of actions: wood-cutting (with an axe), walking, rowing and dancing, were selected from an action database (Vanrie & Verfaillie, 2004). The PLF we labelled and introduced to participants as 'dancer' was originally labelled as 'waving' by the authors of the database, but was described by participants in the original study as 'dancing' (c.f., Vanrie & Verfaillie, 2004). The most important criteria for the action selection were that they involved whole-body movement, i.e., all the dots representing major joints were in motion, and that they could be easily expressed by a single noun. On every trial, PLFs were presented, facing to the left or to the right in sagittal view (90°), in the centre of the RDM aperture. They were shown either in their coherent form, comprising a human figure in action (target present condition), or in a scrambled form, where the initial locations of the landmark dots were randomly positioned within the perimeter of the coherent PLF, while their individual kinematics were preserved. This manipulation renders the target unrecognizable as a coherent biological figure in motion (target absent condition).

On every trial, a coherent or scrambled PLF was embedded in an RDM mask. The masking RDM dots were identical to the target PLF dots in size and luminance, but their kinematics were different. On every trial, a certain proportion of RDM dots (see below) moved coherently in an upward or downward direction, while the rest (i.e., incoherent dots) were re-drawn in a random location at every monitor refresh. Piloting of this masking technique showed that masking efficacy increased with decreasing coherence.

Individual masking levels (i.e., the percentage of RDM dots surrounding the PLF target, moving coherently in an upward or downward direction) were determined for each of the four PLF types and per participant using a Bayesian adaptive staircase procedure (QUEST; Watson & Pelli, 1983). For every action, the threshold was collapsed across upward and downward RDM motion direction as well as across left and right PLF orientation. At the end of the staircase procedure (96 trials per action), we extracted four thresholds for every participant, which reflected the masking level at which the four actions yielded approximately 75% accuracy on a biological motion detection task (see below). Participants who did not reach the 75% accuracy performance on all four actions even when all the noise dots were moving coherently (i.e., at the easiest level of target detection) during the thresholding procedure were excluded from the experiment.

Procedure

Participants were seated in a dimly lit room, approximately 60cm away from the monitor. Stimuli were displayed on an Acer monitor (17”, 1280x1024, 60Hz refresh rate). The participants received both spoken and written instructions (on the screen) prior to doing the experiment.

The experiment consisted of three parts: familiarization, practice and thresholding, and cueing experiment, all described in detail below.

Familiarization: The experiment started with a short familiarization session, during which all the visual targets and lexical cues were presented to the participants, with the instruction to carefully observe the stimuli. The presentation of the PLFs was accompanied by a one sentence description of the type of action they engaged in, e.g., "Je ziet zometeen een figuur, die wandelt" ("*You will see a walking figure*").

Practice and thresholding: The practice session consisted of one block (128 trials) and was shortly followed by the thresholding session of three blocks (128 trials per block). Participants were instructed to monitor the screen and to indicate on each trial whether they detected coherent biological motion or not ("Do you see coherent biological motion, yes or no?"), as quickly and accurately as possible. Participants responded on a button box with left or right index fingers. The trial structure of the thresholding procedure follows that of the cueing experiment, illustrated in **Figure 3.1**, but in order to obtain non-biased estimates of PLF detection, cues were not presented.

Cueing experiment: The cueing part of the experiment consisted of 4 blocks, 128 trials each (512 trials in total), and had the same task and trial structure as the thresholding session (**Figure 3.1**). The only difference was that in this part, participants were presented with either no language cues or lexical cues prior to the visual target, following the design of the experiment. The cues were presented in the centre of the screen (font style: 'lucidatypewriter', font size 18). Trial presentation order was fully randomised.

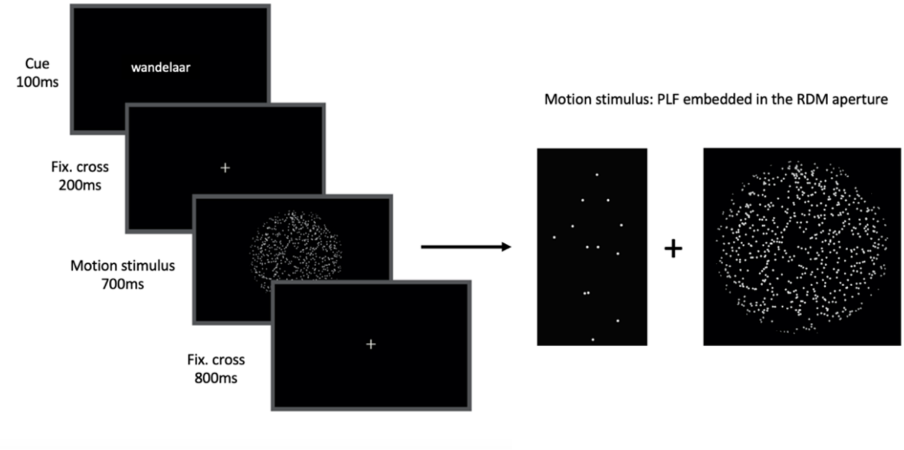


Figure 3.1. Trial design. A string cue is displayed at the beginning of every trial, before the target. The visual motion stimulus, comprised of a PLF figure (coherent or scrambled) embedded in an RDM aperture (upward or downward moving dots at a predetermined level of coherence), is presented in the middle of the screen. Participants have 1.5 sec to respond – 700ms (motion stimulus) + 800ms (fixation cross) – with the instruction to press the button, answering whether they saw coherent biological motion yes or no, as soon as possible.

Analysis

Data analysis was performed on 40 complete datasets. Prior to the analyses, trials with reaction times (RTs) 2.5 SD or more from the grand mean were excluded (trials with RTs above 1345.84ms and below 427.69ms). This resulted in the exclusion of 460 out of 20480 trials (2.2% of trials).

Accuracy and RTs: We were interested in how the lexical cue categories (biological motion, biological form and general motion) influenced both the detection of coherent and rejection of scrambled PLFs. We expected the cues that exert a facilitatory effect on the detection of a coherent target to also be detrimental to the rejection of scrambled targets (i.e., would lead to a higher false alarm rate) and vice versa. First, we compared the three lexical cue categories to the control (no language) cue category, for coherent and scrambled PLF conditions respectively. Further, we aimed to examine differences in the magnitude of the lexical cueing effect as a function of the number of features shared between lexical cue and target, and thus compared the lexical cue categories with one another, again separately for coherent and scrambled PLFs, in a post hoc analysis.

A Bayesian approach allows us to quantify uncertainty in relation to our findings by means of obtaining probability distributions for our parameters of interest rather than a single point estimate, as with frequentist analyses. Further, the three experiments reported here build upon each other, both theoretically and empirically, allowing us to specify priors for each analysis based on the results from the previous experiment. Therefore, we ran Bayesian linear mixed effects models, as implemented in the R package brms (Bürkner, 2017, 2018). Post hoc analyses were conducted with the R package emmeans (Lenth, 2018).

Both accuracy (Bernoulli distribution, logit link) and RT (Gaussian distribution, identity link, with log-transformed RTs) models were fitted with the maximal, hypothesis-driven, non-singular structure supported by the data (Barr et al., 2013; Bates et al., 2018; Matuschek et al., 2017). The resulting model consisted of the predictor ‘cue category’ (4 levels: biological motion, biological form, general motion and no language) nested under the predictor ‘PLF coherence’ (2 levels: coherent, scrambled) as fixed effects, and by-subject and by-item random intercepts and slopes for PLF coherence as random effects. For fixed effects, we used simple effect coding, with the PLF coherence predictor coded as (coherent: 0.5, scrambled: -0.5), and the cue category predictor coded as (no language cues were base coded as -0.25, contrasting condition of each column as 0.75).

Bayesian models were run with informative priors appropriate for the effects expected in this study. Priors for both accuracy and RT models’ contrasts of interest were taken from the estimates and 95% credible intervals of the posteriors of pilot data from three participants. Prior sensitivity analyses were performed to ensure the choice of priors did not influence our results. Both models were fit with 4 chains, 10000 iterations each, 3000 of which were the warm-up phase, with a thinning factor of 3. To assess correct convergence, we verified that there were no divergent transitions, R-hats were all equal to one, the number of effective samples were at least 10% of post-warmup samples, and chains were visually inspected for stationarity.

Signal Detection Analysis (Criterion and d’): In order to separate participants’ conceptual, perceptual and decision biases from perceptual sensitivity/discriminability, we calculated Criterion and d’ scores for our four cue categories per participant. These values were computed based on z-transformed scores of participants' hit rates (correctly reporting the presence of a coherent PLF) and false alarm rates (FA; incorrectly reporting the presence of a coherent PLF when the target was scrambled and therefore non-existent). Both scores were further compared across the four cueing conditions at the group level using unequal variance Bayesian mixed models (Gaussian distribution), with cue category as a fixed effect and by-Subject intercept as a random effect. The fixed effect had the same contrast coding scheme as in the accuracy model, and a lightly regularizing, normally distributed prior with mean of 0 and standard deviation of 2.

Results

Accuracy

For coherent PLFs (target present condition; mean accuracy: 86.68%; see **Figure 3.2.**), the accuracy model revealed higher PLF detection hit rate for biological motion cues, compared to each of the other three cue categories (see **Table 3.1.**). The hit rate advantage for biological motion cues was found to be biggest compared to general motion cues, followed by no language cues, and was smallest compared to biological form cues (see Supplementary **Figure S3.1.**).

For scrambled PLFs (target absent condition; mean accuracy: 89.17%; see **Figure 3.2.**), the model showed the highest false alarm rate (lowest correct rejection) on trials with biological motion cues compared to the other three cue categories respectively (see **Table 3.1.**). This difference was again biggest when biological motion cues were contrasted with general motion cues, followed by no language cues, and smallest compared to biological form cues (see Supplementary **Figure S3.1.**).

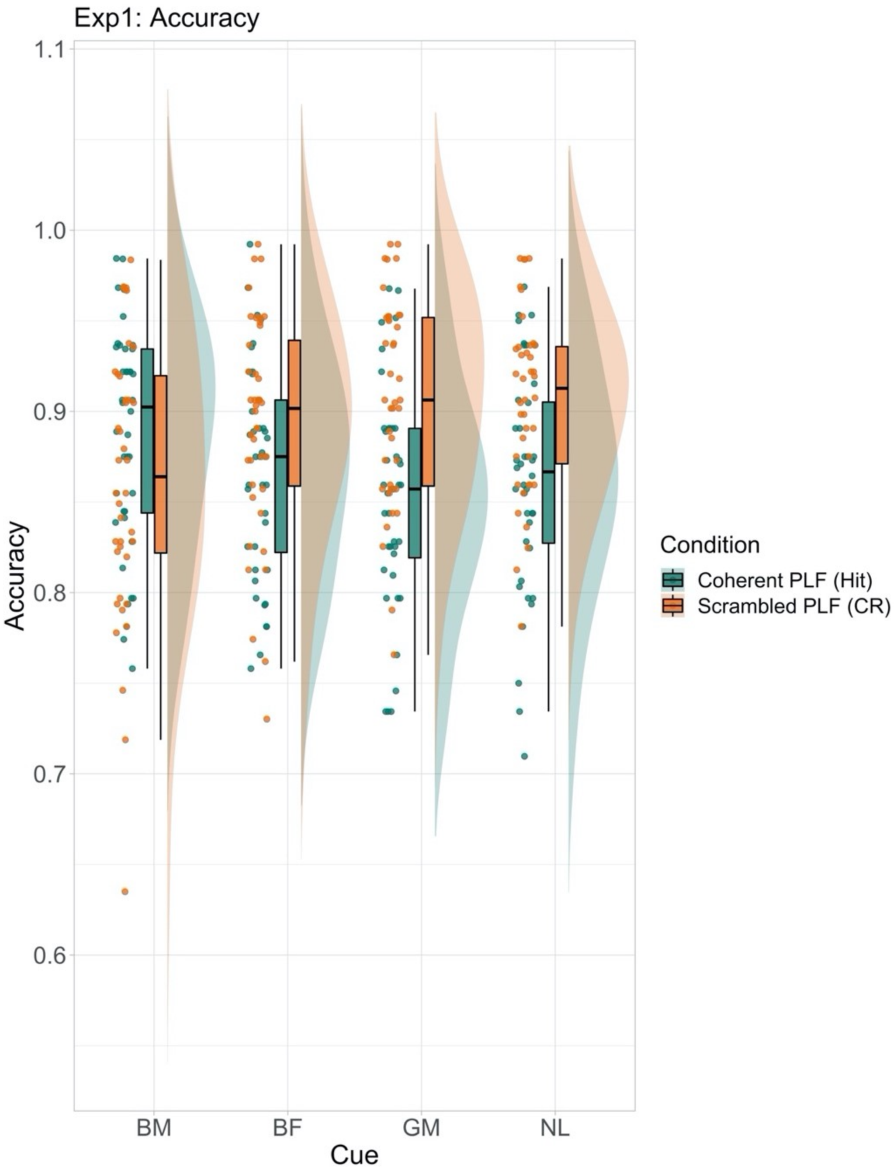


Figure 3.2. Accuracies from Experiment 1. Raw data for coherent (hit) and scrambled (correct rejection; CR) PLF conditions for Experiment 1. Cue names are as follows: BM – biological motion, BF – biological form, GM – general motion, NL – no language.

Criterion and d’

Criterion differed across cue categories as follows: biological motion vs. no language (estimate = -0.17, 95%CrI = -0. 25 - -0.10); biological motion vs. biological form (estimate = -0.12, 95%CrI = -0.20 - -0.04); biological motion vs. general motion (estimate = -0.22, 95%CrI = -0.29 - -0.15); biological form vs. general motion (estimate = -0.1, 95%CrI = -0.17 - -0.03). Sensitivity (d’) did not vary as a function of cue category. These results show that participants were biased towards reporting coherent biological motion (more liberal with their ‘yes’ answers) when cued by biological motion words, compared to the other three cue categories (see Supplementary **Table S3.2.**).

RTs

For coherent PLFs (target present; mean RT: 852.89ms; see **Figure 3.3.**) the model revealed fastest RTs for biological motion cues compared to each of the other three cue categories (see **Table 3.1.**). Furthermore, biological form cues led to faster RTs compared to no language cues. The RTs for biological motion cues were fastest compared to general motion and no language cues, while that difference was smallest compared to biological form cues (see Supplementary **Figure S3.2.**).

For scrambled PLFs (target absent; mean RT: 896.86ms; see **Figure 3.3.**), the model showed slowest RTs for biological motion cues compared to the general motion cues (see **Table 3.1.**). Furthermore, both biological form and general motion cues led to faster RTs compared to no language cues (see Supplementary **Fig**. **S3.2.**).

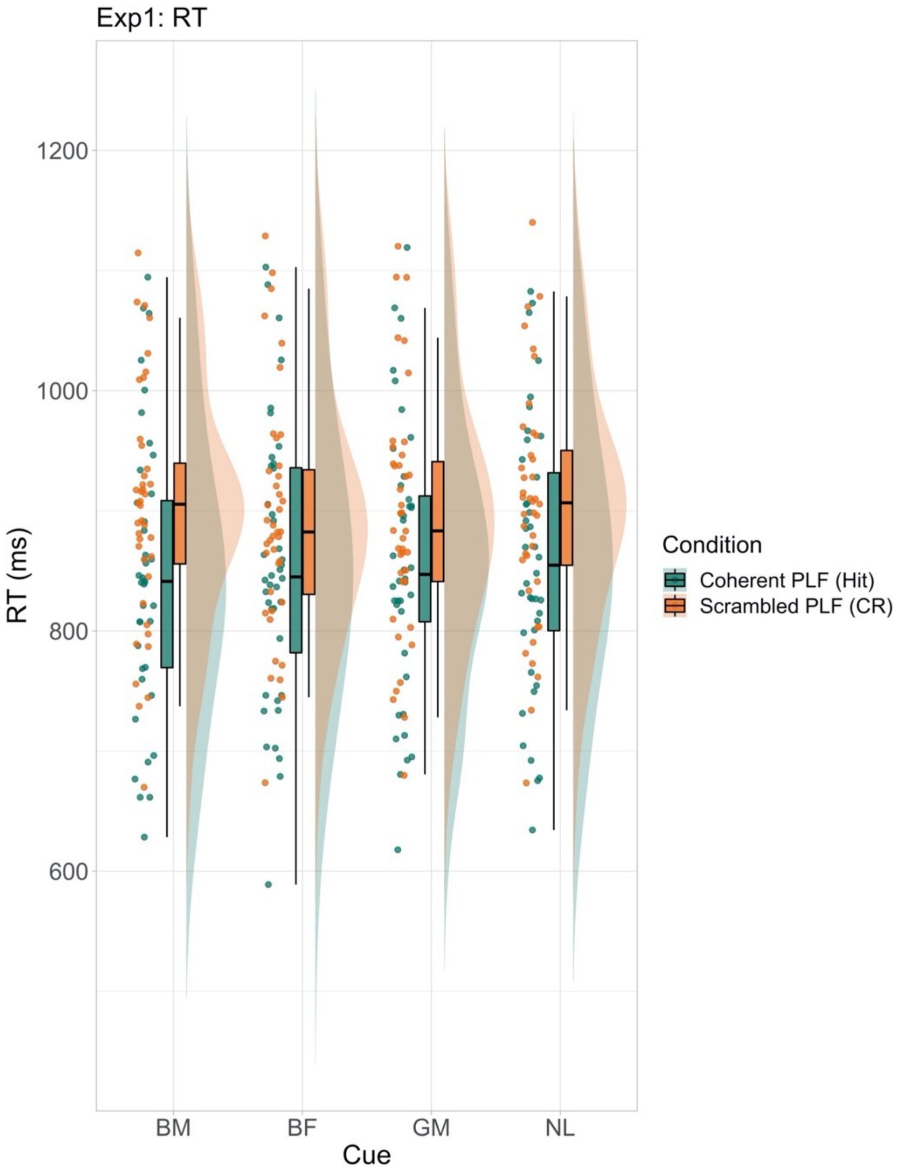


Figure 3.3. RTs from Experiment 1. Raw data for coherent and PLF conditions for Experiment 1. Cue names are as follows: BM – biological motion, BF – biological form, GM – general motion, NL – no language.

Table 3.1.
Accuracy and RT results for Experiment 1

	Accuracy Coherent PLF	Accuracy Scrambled PLF	log(RT) Coherent PLF	log(RT) Scrambled PLF
Contrasts	*Estimate (95% CrI)*	*Estimate (95% CrI)*	*Estimate (95% CrI)*	*Estimate (95% CrI)*
BM vs. NL	**0.23** **(0.06 – 0.41)**	**-0.44** **(-0.62 – -0.26)**	**-0.02** **(-0.03 – -0.01)**	-0.00 (-0.01 – 0.01)
BF vs. NL	0.05 (-0.11 – 0.22)	-0.15 (-0.34 – 0.04)	**-0.01** **(-0.02 – -0.00)**	**-0.01** **(-0.02 – -0.00)**
GM vs. NL	-0.09 (-0.26 – 0.07)	0.04 (-0.16 – 0.23)	-0.00 (-0.01 – 0.01)	**-0.01** **(-0.02 – -0.00)**
BM vs. BF	**0.18** **(0.01 – 0.35)**	**-0.29** **(-0.46 – -0.12)**	**-0.01** **(-0.02 – -0.00)**	0.01 (-0.00 – 0.02)
BM vs. GM	**0.33** **(0.16 – 0.50)**	**-0.48** **(-0.66 – -0.29)**	**-0.02** **(-0.03 – -0.01)**	**0.01** **(0.00 – 0.02)**
BF vs. GM	0.15 (-0.02 – 0.31)	-0.19 (-0.38 – 0.01)	-0.01 (-0.02 – 0.00)	0.00 (-0.01 – 0.01)

Note: Cue names are as follows: BM – biological motion, BF – biological form, GM – general motion, NL – no language. Bolded are estimates and credible intervals that did not cross zero.

3.2.2. Experiment 2

Experiment 1 showed that cues encoding multiple biological motion features enhanced coherent PLF detection and interfered with scrambled PLF correct rejection. However, all biological motion trials included congruent cue-target pairs. We thus cannot rule out that the observed effect was due to this cue-target contingency. The finding that congruent cues facilitate coherent PLF detection also raises the question of whether incongruent cues (i.e., cues with both form and kinematics features encoded but not congruent with the PLF target) may lead to equally strong detrimental effects. We therefore extended our investigation by incorporating incongruent biological motion, with the human form feature congruent and kinematics feature incongruent with the target, cue-target pairs into the paradigm.

Participants

Fifty-five native Dutch speakers (45 female, 10 male, mean age: 23.56, age range: 19-33) recruited from the MPI participant database took part in the experiment. Fifteen participants failed to reach the inclusion criterion during the thresholding procedure and were excluded from the analysis, resulting in 40 complete datasets (31 female, 9 male, mean age: 23.58, age range: 19-33). All participants were right-handed and had normal or corrected-to-normal vision, and no reading difficulties. All the participants gave their informed consent and received monetary compensation for their participation.

Stimuli

The same stimuli and procedure were used as in Experiment 1, with the following changes. The four PLFs were slightly rotated from 90 to 45°, left and right facing profile, to create a more visible form angle, i.e., to minimize the potential crossing of the landmark dots. For linguistic cues, the biological form cue category was replaced with incongruent biological motion cues (e.g., ‘dancer’ followed by the ‘walker’ target), in order to test the effect of congruence within the biological motion category on PLF detection. Every cue was paired with each PLF stimulus, resulting in two incongruent pairings (general motion and incongruent biological motion) and one congruent pairing (congruent biological motion).

Procedure

The experiment had the same parts as Experiment 1, but this time after the familiarization section, participants went through 4 practice blocks before proceeding to the thresholding procedure. An additional practice block was added to increase participant familiarity with the stimuli, because mean accuracy in Experiment 1 was well above the level that had been estimated during the thresholding procedure (75%), suggesting that some learning may have taken place after the practice block and during the thresholding procedure. Stimulus presentation order in all parts of the experiment was again fully randomised.

Analysis

Data analysis was performed on 40 complete datasets. Prior to the analyses, trials with RTs 2.5 SD or higher from the grand mean were excluded (trials with RTs above 1367.145 and below 441.7752). This resulted in the removal of 528 out of 20480 trials (2.5% of trials).

Bayesian linear mixed effects models for the **accuracy, RT, Criterion and d' analyses** had the same fixed/random effects structure and coding scheme as Experiment 1. Priors (Gaussian distribution) for both accuracy and RT models contrasts of interest were taken from the posterior estimates and 95% credible intervals from Experiment 1 and differed across conditions in line with the estimates and credible intervals in Experiment 1. Priors for Criterion and d' were again lightly regularizing, as in Experiment 1.

Results

Accuracy

For coherent PLFs (target present; mean accuracy: 80.78%; see **Figure 3.4.**), the model revealed a higher PLF detection hit rate for congruent biological motion cues compared to the other three cue categories respectively (see **Table 3.2.**). This difference was larger compared to semantically invalid (general motion and incongruent biological motion) cues than for no language cues (see Supplementary **Figure S3.3.**).

For scrambled PLFs (target absent; mean accuracy: 89.17%; see **Figure 3.4.**), the model showed the highest false alarm rate (lowest correct rejections) for biological motion cues compared the two lexical cue categories respectively (see Supplementary **Figure S3.3.**).

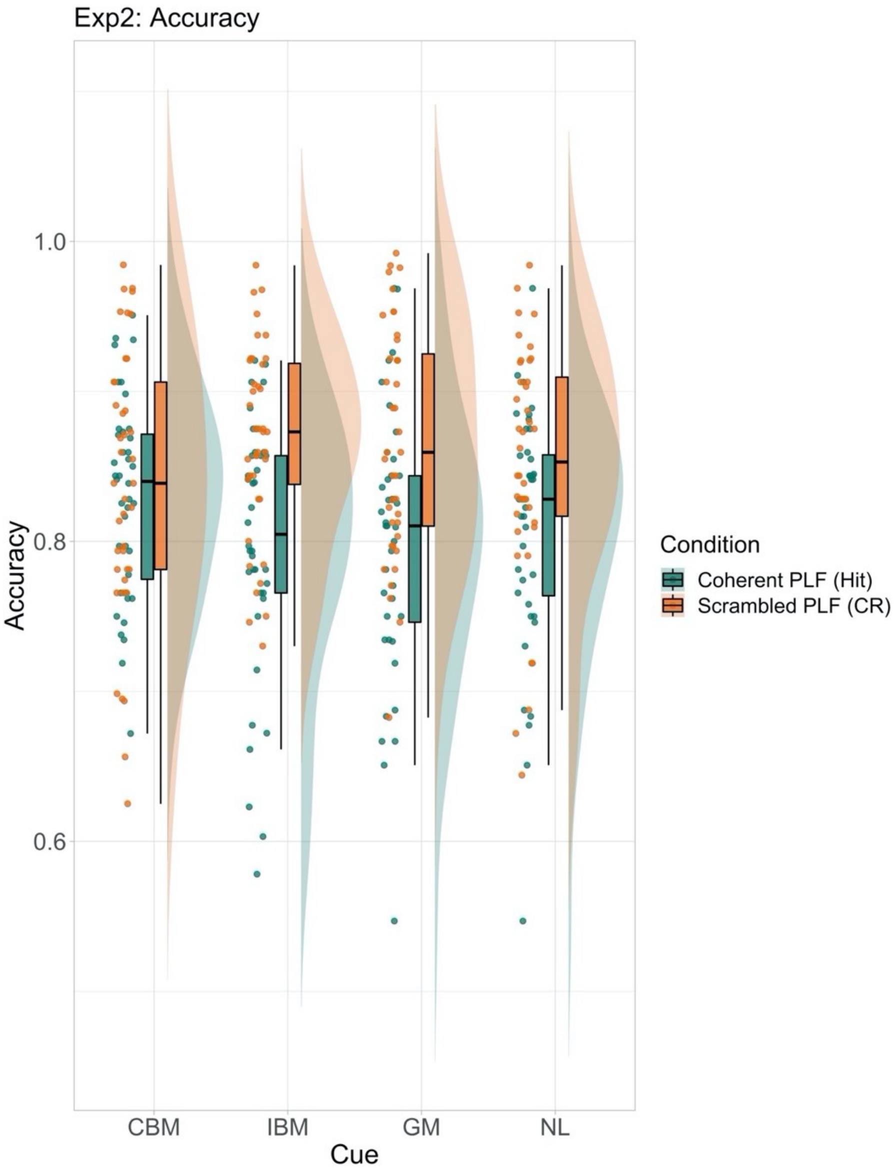


Figure 3.4. Accuracies from Experiment 2. Raw data for coherent (hit) and scrambled (correct rejection; CR) PLF conditions for Experiment 2. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, GM – general motion, NL – no language.

Criterion and d’

Criterion differed as a function of cue category as follows: congruent biological motion vs. general motion (estimate = -0.12, 95%CrI = -0.2 - -0.04); congruent biological motion vs. incongruent biological motion (estimate = -0.11, 95%CrI = -0.18 - -0.03). Sensitivity (d’) did not vary as a function of cue category. The results indicate that participants were less conservative (less likely to report target as absent) with their answers when cued by congruent biological motion words than other lexical cues (see Supplementary **Table S3.2.**).

RTs

For coherent PLFs (target present; mean RT: 873.97ms; see **Figure 3.5.**) the model revealed fastest RTs for congruent biological motion cues compared to other three cue categories respectively (**Table 3.2.**). Furthermore, general motion cues led to faster RTs than incongruent biological motion cues, suggesting that incongruent biological motion cues were most detrimental to the task (see Supplementary **Figure S3.4.**).

For scrambled PLFs (target absent; mean RT: 908.31ms, see **Figure 3.5.**), the model did not reveal any notable differences between cue categories (see Supplementary **Figure S3.4.**).

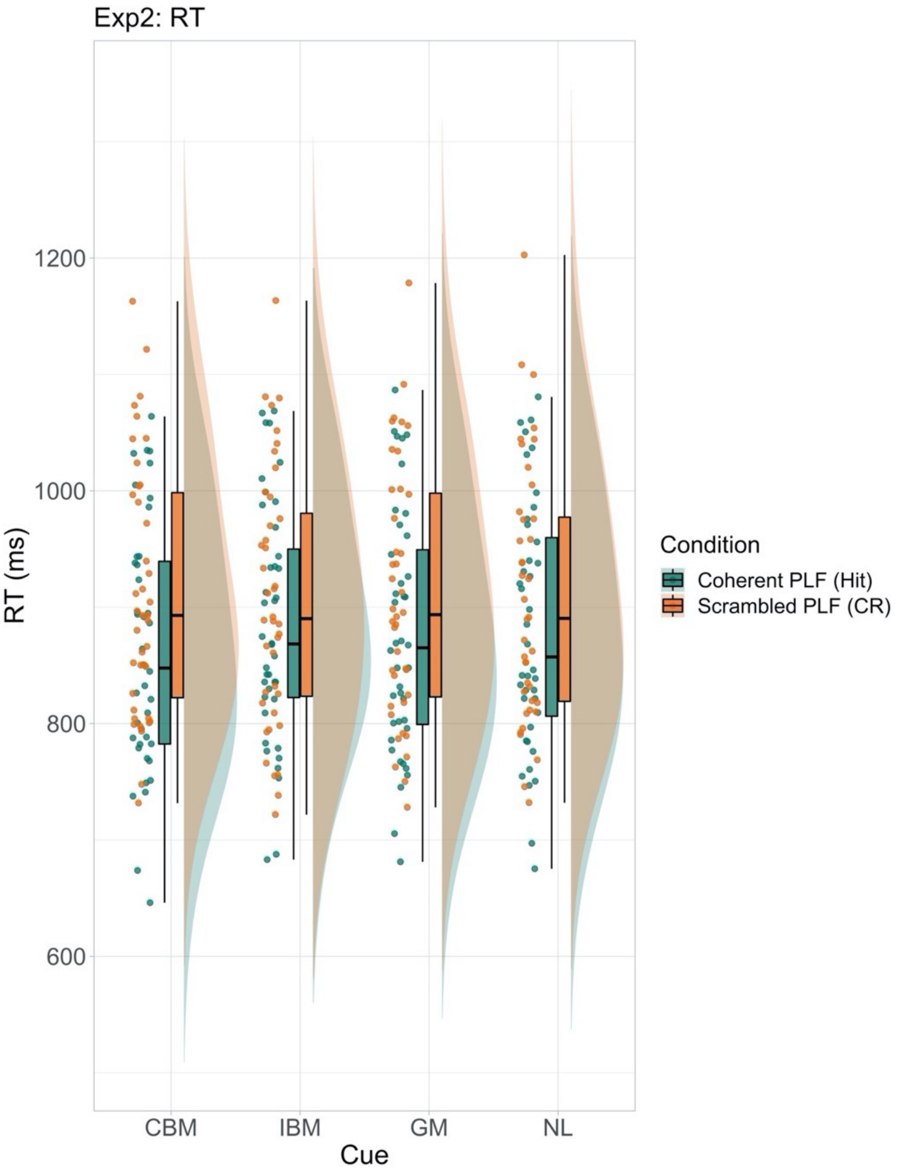


Figure 3.5. RTs from Experiment 2. Raw data for coherent and PLF conditions for Experiment 2. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, GM – general motion, NL – no language.

Table 3.2.
Accuracy and RT results for Experiment 2

	Accuracy Coherent PLF	Accuracy Scrambled PLF	log(RT) Coherent PLF	log(RT) ScrambledPLF
Contrasts	*Estimate (95% CrI)*	*Estimate (95% CrI)*	*Estimate (95% CrI)*	*Estimate (95% CrI)*
CBM vs. NL	**0.15** **(0.01 – 0.30)**	-0.12 (-0.28 – 0.03)	**-0.02** **(-0.03 – -0.01)**	0.00 (-0.01 – 0.01)
GM vs. NL	-0.07 (-0.21 – 0.07)	0.13 (-0.03 – 0.28)	-0.01 (-0.01 – 0.00)	-0.00 (-0.01 – 0.01)
IBM vs. NL	-0.06 (-0.21 – 0.08)	0.13 (-0.02 – 0.29)	0.00 (-0.00 – 0.01)	-0.00 (-0.01 – 0.01)
CBM vs. GM	**0.22** **(0.08 – 0.37)**	**-0.25** **(-0.41 – -0.09)**	**-0.02** **(-0.03 – -0.01)**	0.00 (-0.01 – 0.01)
CBM vs. IBM	**0.22** **(0.07 – 0.36)**	**-0.26** **(-0.42 – -0.1)**	**-0.03** **(-0.04 – -0.02)**	0.01 (-0.00 – 0.01)
GM vs. IBM	-0.01 (-0.15 – 0.13)	-0.01 (-0.17 – 0.16)	**-0.01** **(-0.02 – -0.00)**	0.00 (-0.01 – 0.01)

Note: Cue names are as follows: CBM – congruent biological motion, GM – general motion, IBM – incongruent biological motion, NL – no language. Bolded are estimates and credible intervals that did not cross zero.

3.2.3. Experiment 3

Experiments 1 and 2 showed that lexical cues encoding multiple biological motion features (both form and kinematics) had the strongest influence on biological motion detection. However, feature encoding in the previous two experiments was manipulated only for lexical cues, while PLF targets always contained both form and kinematics features. We therefore could not dismiss the possibility that while both form and kinematics features do need to be represented by linguistic labels, form feature alone conveyed by the visual target might be enough for the lexical cueing effect on perception to be observed.

In Experiment 3, we aimed to test whether lexical cues have an effect only on the final target form, arising from the point-light dots configuration, or if they are instead guiding the compositional process reliant on multiple diagnostic features (i.e., kinematics of the dots as well as their placement) of biological motion. For that reason, in addition to the naturally moving PLFs, we included a condition with visually absent biological kinematics, but present biological form (PLFs frozen in a canonical stance and moving horizontally in space; 'gliders').

Participants

Sixty native Dutch speakers (47 female, 13 male, mean age: 23.37, age range: 20-33) recruited from the MPI participant database took part in the experiment. Twenty participants failed to reach the inclusion criterion during the thresholding procedure and were therefore excluded from the analysis, resulting in 40 complete datasets (32 female, 8 male, mean age: 23.5, age range: 20-33). All participants were right-handed and had normal or corrected-to-normal vision, and no reading difficulties. All participants gave their informed consent and received monetary compensation for their participation.

Stimuli

Experiment 3 had the same experiment and trial structure as Experiment 2, with some modifications. Lexical cues remained the same as in Experiment 2, but along with the naturally moving PLFs ('naturals'; presented in Experiment 1 and 2) an additional type of motion – gliding PLFs ('gliders', speed: 0.528mm/frame) – was introduced as a visual target on 50% of the trials in an event related manner. Gliders were PLF animations captured in a recognizable frame and made to move rigidly back and forth in space inside of the RDM aperture, at the speed that approximately matched the speed of the RDM dots. This manipulation made it possible for the PLFs to retain form in the absence of local kinematics, thus allowing for evaluation of the cueing magnitude of the form-only visual target.

The addition of gliders required a change of task: the figures were presented in either their upright (normal) form or inverted (upside down) form. Inverting the PLFs was expected to impair the successful composition of the dots into a resolved, canonical human form, and thereby to have a similar effect on biological motion detection as the scrambling in our Experiments 1 and 2(Pavlova & Sokolov, 2000). Participants were therefore asked to discriminate between upright and inverted PLFs, placing the focus on the biological form rather than the motion itself, and making the task therefore less directly related to the lexical cues.

Procedure

Similar to the previous two experiments, Experiment 3 started with a short familiarization section, succeeded by practice and thresholding. This comprised two practice blocks and four thresholding blocks, 128 fully randomised trials each. Trials had the same structure as the previous two experiments. Thresholding was carried out individually for naturally moving and gliding PLFs, resulting in 8 thresholds per participant extracted after 64 trials per action and motion type, but was otherwise identical to the previous two experiments. The cueing experiment consisted of 4 blocks, 256 fully randomised trials each (1024 trials in total, 512 trials per motion type).

Analysis

Forty complete datasets were analysed. Prior to the analysis, trials with RTs 2.5 SD or higher from the grand mean were excluded (trials with RTs above 1339.734 and below 350.2306). This resulted in the removal of 1519 out of 40960 trials (3.7% of trials excluded).

For accuracies, RTs, Criterion and d', the models had a different fixed and random effects structure from Experiment 1 and 2 due to the extra 'PLF motion type' factor (gliding or naturally moving PLFs, labelled gliders and naturals, respectively). For accuracy and RTs, the fixed effects structure therefore included the predictor 'cue category' (congruent biological motion, general motion, incongruent biological motion and no language) nested under 'PLF orientation' predictor (upright, inverted) nested under 'PLF motion type' predictor (naturals, gliders). This structure allowed evaluation of the effect of lexical cues for each motion type and PLF orientation type. Random effects structure consisted of by-subject and by-item random intercepts and slopes for PLF motion and PLF coherence but not their interaction. For fixed effects, the PLF motion predictor was coded as (naturals: 0.5, gliders: -0.5), and the PLF coherence and cue category predictors was coded as in Experiment 2. Priors for both accuracy and RT models were taken from the posterior estimates and 95% credible intervals from Experiment 2, while signal detection indices (Criterion and d’) were analysed for naturals and gliders respectively in the same way as in the Experiment 2.

Results

Both accuracy and RT analyses (mean accuracy: 90.10%, see **Figure 3.6.** for accuracy; mean RT: 806.84ms; see **Figure 3.7.** for RT) showed no discernible effects for any of the cueing categories on the discrimination of the gliding PLFs**.**

For naturally moving PLFs (‘naturals’, mean accuracy: 85.46%), accuracy results showed a higher false alarm rate for inverted PLFs when cued by congruent biological motion cues compared to the incongruent ones (estimate = -0.15, 95%CrI = -0.29 - -0.01; see Supplementary **Figure S3.5.**). The models revealed no differences for either Criterion or d’ as a function of cue categories (see Supplementary **Table S3.3.**).

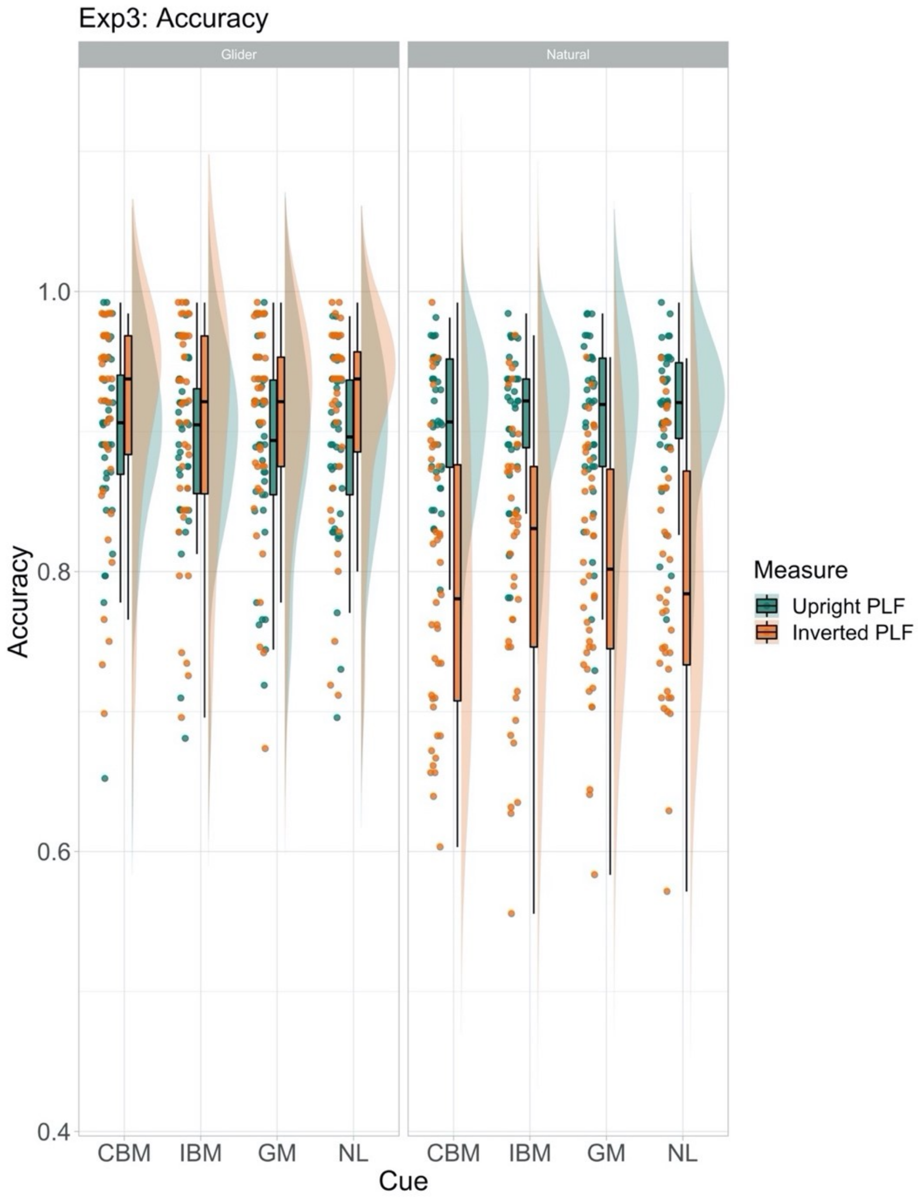


Figure 3.6. Accuracies from Experiment 3. Raw data for glider and natural PLF motion conditions, for upright and inverted PLF orientation conditions for Experiment 3. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, GM – general motion, NL – no language.

RT results (‘naturals’; mean RT: 859.81ms) revealed a faster detection rate for naturally moving upright PLFs when cued by congruent biological motion compared to incongruent cues (estimate = -0.01, 95%CrI = -0.02 - -0.002; see Supplementary **Figure S3.6.**).

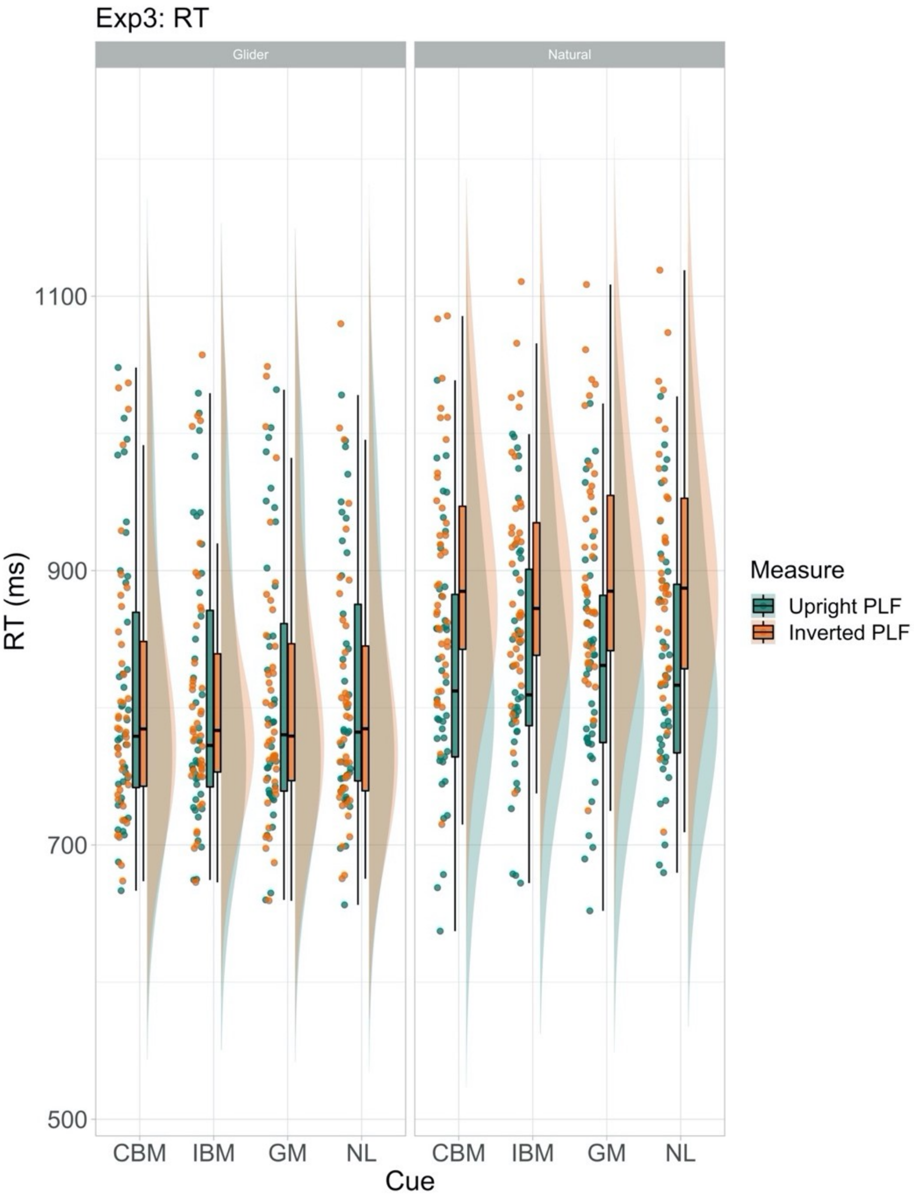


Figure 3.7. RTs from Experiment 3. Raw data for glider and natural PLF motion conditions, for upright and inverted PLF orientation conditions for Experiment 3. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, GM – general motion, NL – no language.

3.3. Discussion

The results presented here show that perception of point-light figures is susceptible to lexical influence. Throughout the three experiments reported here, our results consistently show that lexical cues do not increase perceptual discriminability between coherent and scrambled PLFs (as reflected in the d' scores), but rather bias the PLF detection towards the cued category (as reflected in the Criterion scores). Namely, we find that our congruent cues lead to more accurate and/or faster detection of coherent PLFs but also slow down and/or impede the correct rejection (higher FA rate) of scrambled PLFs (details below). In line with the arguments posited by the label-feedback hypothesis and our understanding of the mechanism underlying such hypothesis described in the Introduction, this bias was interpreted as mainly conceptual and perceptual in nature. While we do not entirely exclude the possibility that decision making processes might have played a small role in the task performance, our experimental design makes it unlikely that they were the driving mechanism behind our results. In all three experiments, the probability of any cue-target pairing appearance was strictly controlled for, the cue-target manipulation was based on the overlap in features, i.e., it was at the level of the semantic content that was not apparent to the participants and even detrimental to their performance, and the response window was very limited. These factors minimised the chance that some kind of post-perceptual, performance-optimising strategy could have been employed during the task performance. Additionally, we found no evidence of a button press bias towards coherent PLF response (‘yes’ answer): neither were ‘yes’ responses given more frequently throughout the experiment, nor were the RTs of FAs shorter than those of CRs/hits. This allowed us to exclude the possibility that the lexically induced bias happened at the response level (e.g., due to motor activation of the finger consistent with the ‘yes’ response) and further confirmed our interpretation of the observed bias as happening at the conceptual and perceptual level.

In Experiment 1, the cueing effect on biological motion detection increased gradually with the number of features encoded in and overlapping between cue and target. In Experiment 2, we observed the same pattern of results as in Experiment 1 and further found that incongruent cues lead to interference with PLF detection. In Experiment 3, biological motion cues were found to only influence PLF orientation discrimination when the PLF had both biological kinematics and form encoded, i.e., with naturally moving PLFs. This finding confirmed that the availability of the form-feature alone encoded in the visual target is insufficient for the lexical cueing effect in this study to be observed.

We found linguistic influence on biological motion perception across tasks, both in biological motion detection and figure orientation discrimination. Figure orientation discrimination does not require the detection or integration of biological kinematics features, but rather puts emphasis on the figure outline, making it a task indirectly related to the kinematics feature encoding. Therefore, these findings suggest that lexically mediated action-relevant features are conceptually and perceptually activated even when the task itself does not directly require their involvement. This in turn supports the claim that lexically mediated activation of category-relevant features occurs automatically, regardless of task (Lupyan & Thompson-Schill, 2012). However, the cueing effect was weaker in the orientation discrimination task, indicating that the perceptual task directly related to the diagnostic features of the stimuli enhances the cueing effect (see Gilbert & Sigman, 2007; Huettig et al., 2020; Kerzel et al., 2009; Maier & Abdel Rahman, 2019).

The compositional nature of the perception of biological motion in this study allowed us to examine how an overlap in the number of features expressed by labels and visual targets affects the strength of a lexical cueing effect on perception. This is particularly relevant because the uniquely powerful influence of lexical cues on perception has been attributed to their ability to activate the neurons coding for the diagnostic features of a labelled category, thereby biasing the perception of visual stimuli towards that category (Barsalou et al., 2003; Edmiston & Lupyan, 2015). Our findings extend the evidence that feature activation plays an important role in the mechanism driving language-perception interaction by showing that the magnitude of both facilitatory when congruent and detrimental when incongruent cueing effects grows with the number of diagnostic features encoded in the lexical cue and the visual target. This argument is supported by the finding that even underspecified and seemingly unrelated, biological form cues, overlapping with the target on a single (form) feature, affected the speed of the PLF detection in Experiment 1, albeit less strongly than cues with multiple feature overlap (i.e., biological motion words), showing that lexical cueing is not an all-or-nothing phenomenon.

Further, the findings in Experiment 3 show that only when the kinematics feature was encoded in the visual stimuli (i.e., naturally moving PLFs), was target perception susceptible to the influence of motion labels. Conversely, when this information was removed from the visual targets (i.e., gliding figures), thus placing the emphasis on the outline (form) alone, the cueing effect disappeared, even though the discrimination performance was similar to that with naturally moving PLFs. This finding suggests that while lexical effects on perception can be observed even with underspecified (single-feature overlap) lexical cues (see above), they cannot survive impoverished visual target stripped of one of the category-defining features.

Existing work indicates that labels can penetrate the perceptual system and modulate its activation (Boutonnet & Lupyan, 2015; Meteyard et al., 2007; Ostarek & Huettig, 2017). These studies, however, have not looked at different aspects of perceptual processing as corresponding to the diagnostic features encoded by the labels. The current study makes that possible by using *compositional* visual targets, i.e., the form of a PLF target is an emergent, global feature, the detection of which is conditional upon a successful local, kinematics-driven configuration(George Mather et al., 1992; Thornton et al., 1998). We found that biological form cues, targeting the form encoding aspect of PLF detection, exerted a weaker top-down influence on biological motion perception than biological motion cues, targeting both the kinematics and form encoding aspects of PLF detection. This finding suggests that the combined lexical encoding of form and kinematic features biases perceptual activation towards the cued category more comprehensively than form features alone. This in turn illustrates that when it comes to lexical influence on perception, the final percept is guided by interactive yet distinct, feature encoding aspects of that influence.

Importantly, we show that the perceptual bias evoked by linguistic labels is robust enough to mislead participants into reporting an erroneous percept, thus supporting previous findings of detrimental effects of incongruent linguistic cues on visual perception (Meteyard et al., 2007). In our study, this effect was exemplified by the finding that scrambled PLFs were perceived as coherent when cued by congruent biological motion cues, whereas incongruent cues impaired coherent PLF detection. These findings suggest that linguistic labels are strong enough to activate a misleading visual template for PLF configuration, leading to an incorrect report when the lexically labelled action does not match that performed by the PLF. This finding is particularly compelling given that we ensured that our participants detected the PLF stimuli with very high accuracy (>75%; anything above 55.1% would be significantly above chance level), showing that the perception of even highly detectable targets is still susceptible to lexical cueing.

In conclusion, this study furthers our understanding of language-perception interaction by empirically attesting feature encoding and overlap as the driving mechanism behind linguistic influence on perception, specifically a lexical cueing effect on biological motion perception. We show that linguistic influence can bias conceptual and perceptual processing towards the diagnostic features of the category conveyed through the linguistic input, even at high levels of detection accuracy. Crucially, we show that this linguistically mediated perceptual pre-activation of category-diagnostic features occurs in an involuntary (i.e., automatic) fashion, irrespective of task demands, whether or not the linguistic information is ultimately beneficial to task performance.

CHAPTER 3

Supplementary Material

Table S3.1.
Lexical cues

Biological form	**Biological motion**	**General motion**
vader *(father)*	houthakker *(woodcutter)*	hagel *(hail)*
broer *(brother)*	wandelaar *(walker)*	sneeuw *(snow)*
oom *(uncle)*	roeier *(rower)*	rook *(smoke)*
echtgenoot *(husband)*	danser *(dancer)*	stoom *(steam)*

The cues were presented in Dutch. English translation is provided in the brackets. The baseline, non-linguistic cue was '####'. Biological motion cues in Experiment 1 were always congruent with the PLF target. In Experiment 2 and 3, they were either congruent or incongruent with the target. General motion cues were always directionally congruent with the RDM mask.

Table S3.2.
D' and Criterion scores for Experiment 1 and 2

Experiment 1			**Experiment 2**		
Cue categories	**d'**	**Criterion**	**Cue categories**	**d'**	**Criterion**
Biological motion	2.45	-0.05	Congruent biological motion	2.06	0.04
Biological form	2.49	0.22	Incongruent biological motion	2.04	0.15
General motion	2.51	0.23	General motion	2.10	0.16
No language	2.52	0.21	No language	2.02	0.09

Table S3.3.
D’ and Criterion scores for Experiment 3 (naturals and gliders)

	Naturals		**Gliders**	
Cue categories	**d’**	**Criterion**	**d’**	**Criterion**
Congruent biological motion	2.27	-0.26	2.86	0.06
Incongruent biological motion	2.32	-0.23	2.77	0.08
General motion	2.29	-0.25	2.76	0.05
No language	2.32	-0.27	2.78	0.09

Posterior densities for the accuracy and RT models in Experiment 1, 2 and 3:

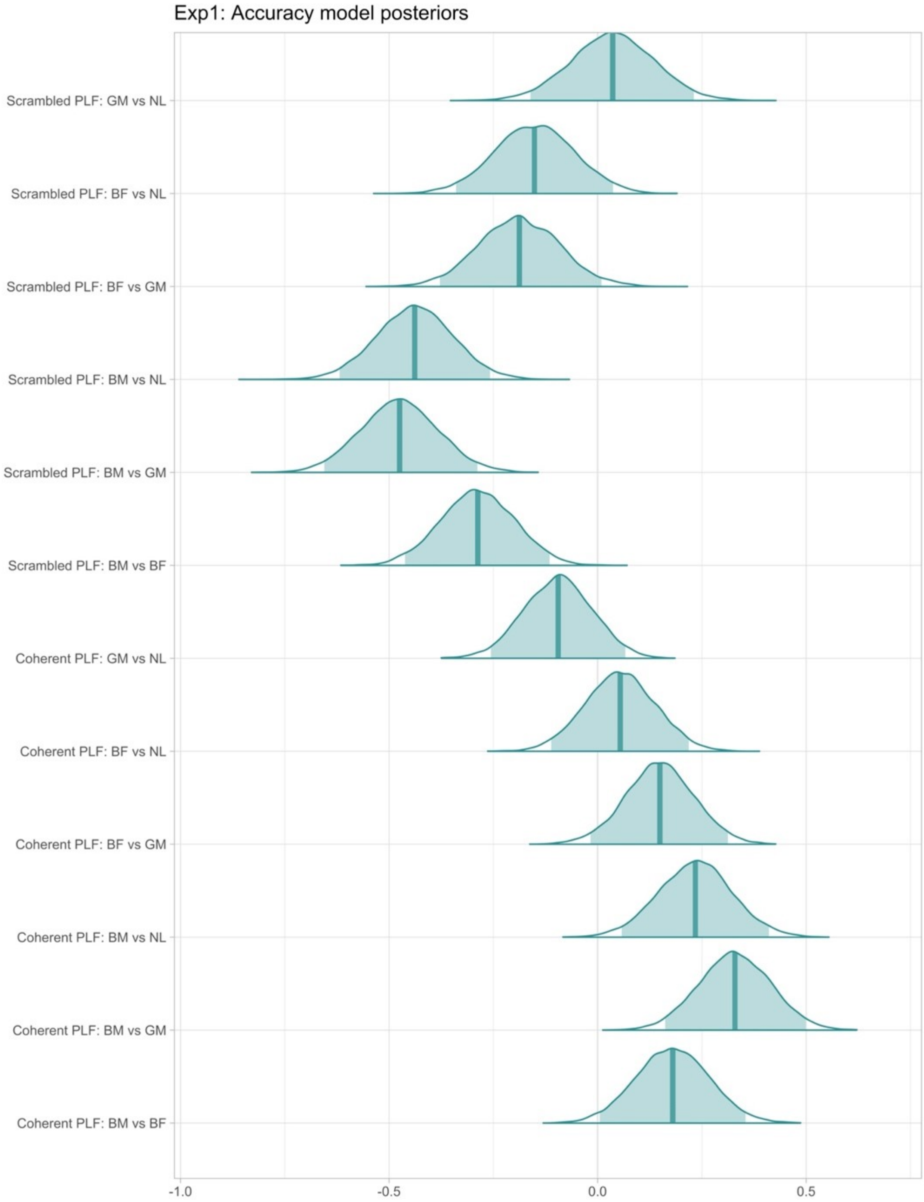


Figure S3.1. Posterior densities for the accuracy model for Experiment 1. Areas filled in green mark 95% credible intervals. Vertical mark is the mean of the distribution. Cue names are as follows: BM – biological motion, BF – biological form, GM – general motion, NL – no language.

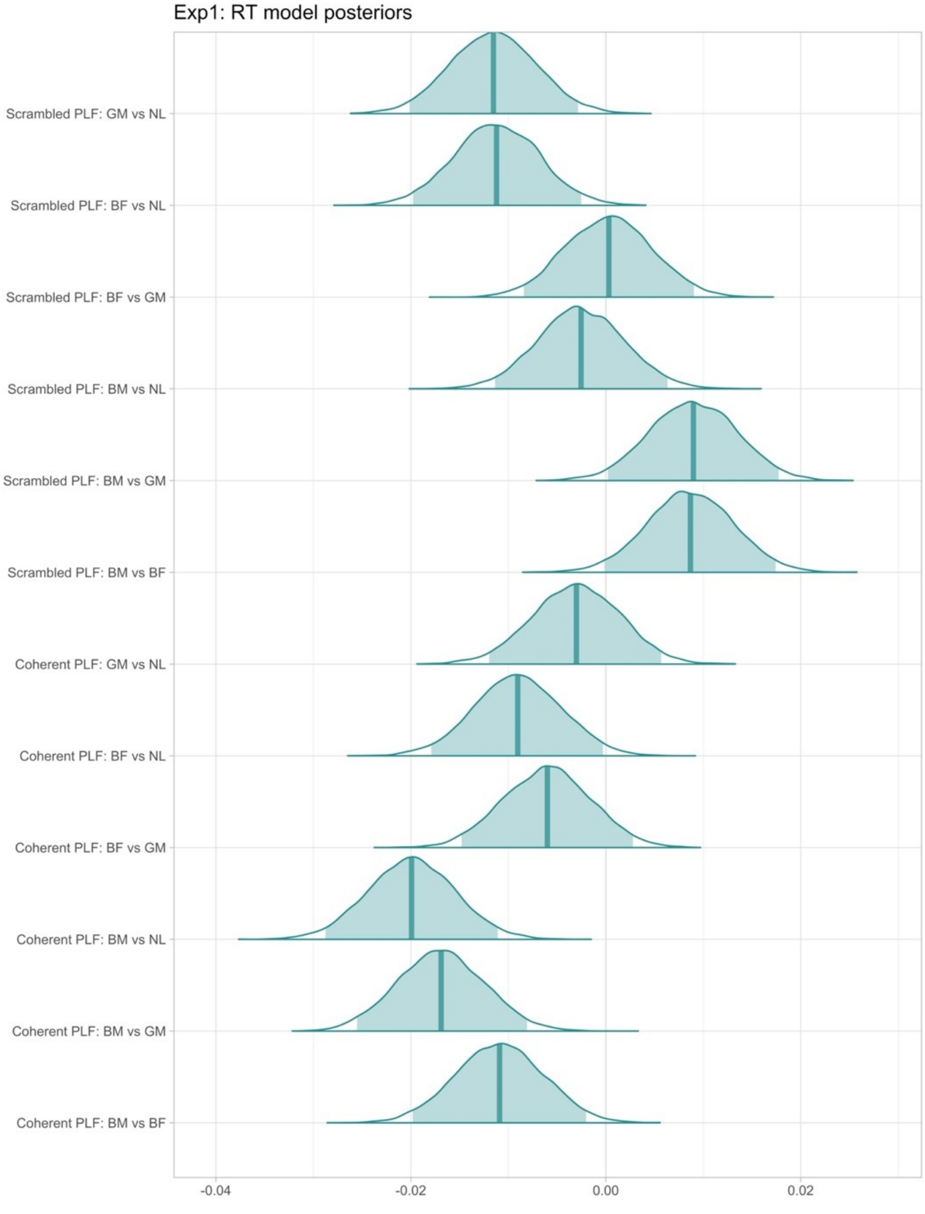


Figure S3.2. Posterior densities for the RT model for Experiment 1. Areas filled in green mark 95% credible intervals. Vertical mark is the mean of the distribution. Cue names are as follows: BM – biological motion, BF – biological form, GM – general motion, NL – no language.

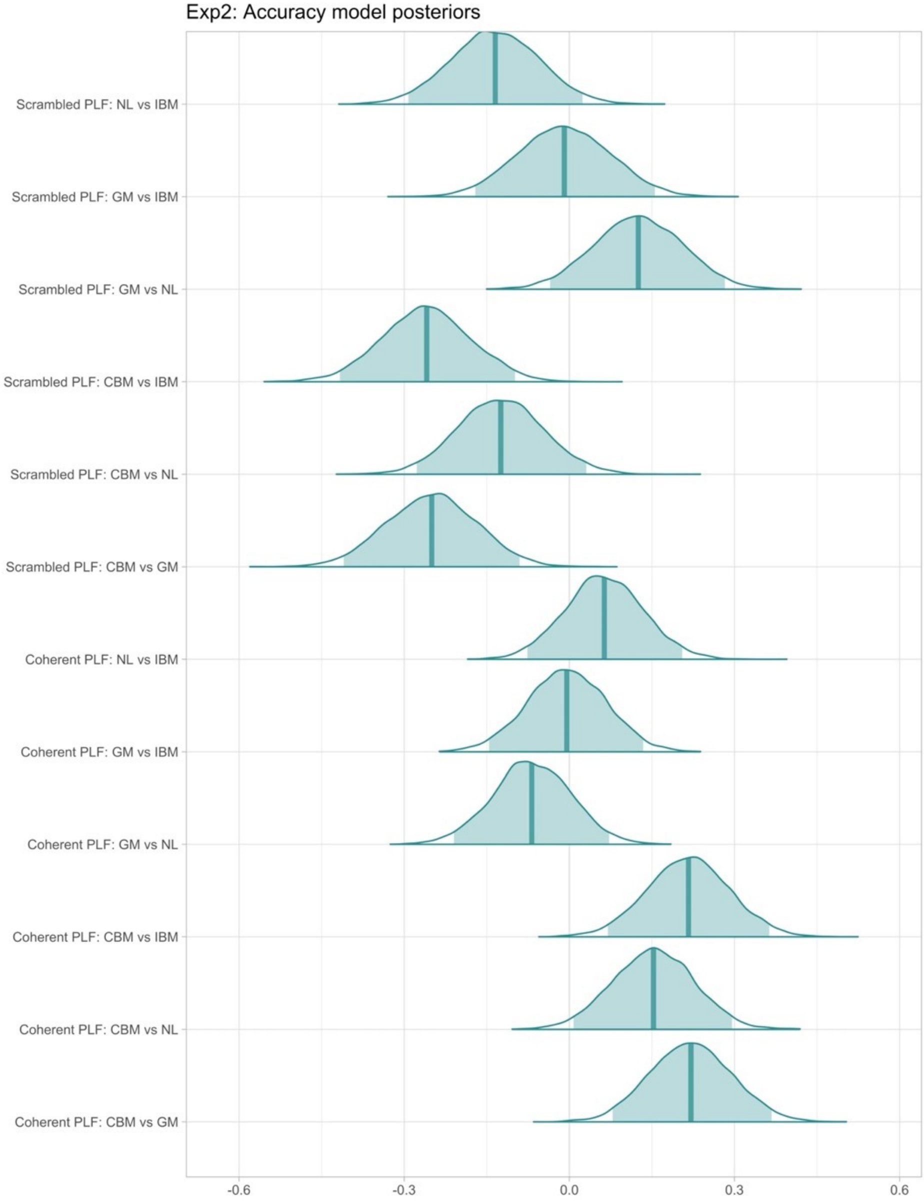


Figure S3.3. Posterior densities for the accuracy model for Experiment 2. Areas filled in green mark 95% credible intervals. Vertical mark is the mean of the distribution. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, GM – general motion, NL – no language.

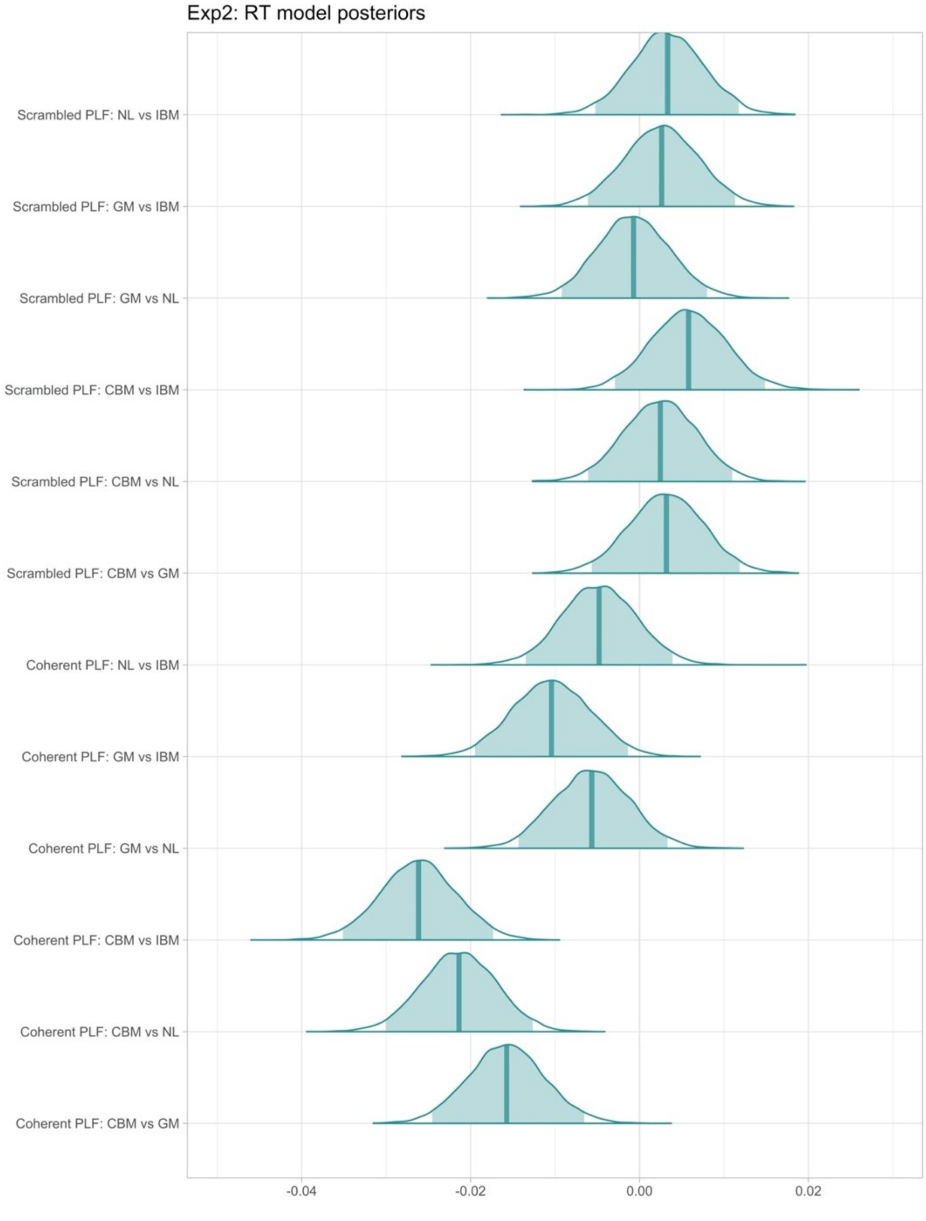


Figure S3.4. Posterior densities for the RT model for Experiment 2. Areas filled in green mark 95% credible intervals. Vertical mark is the mean of the distribution. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, GM – general motion, NL – no language.

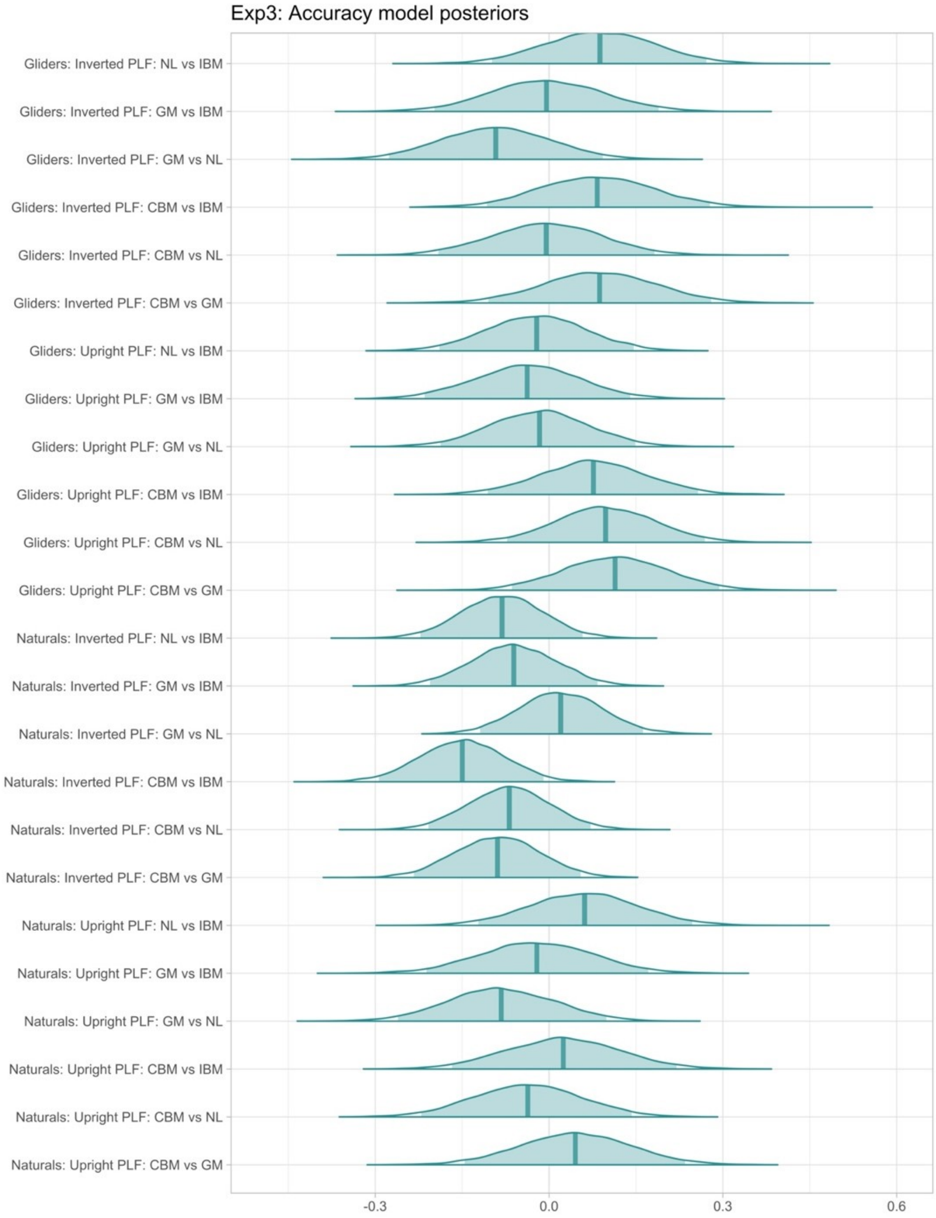


Figure S3.5. Posterior densities for the RT model for Experiment 3. Areas filled in green mark 95% credible intervals. Vertical mark is the mean of the distribution. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, GM – general motion, NL – no language.

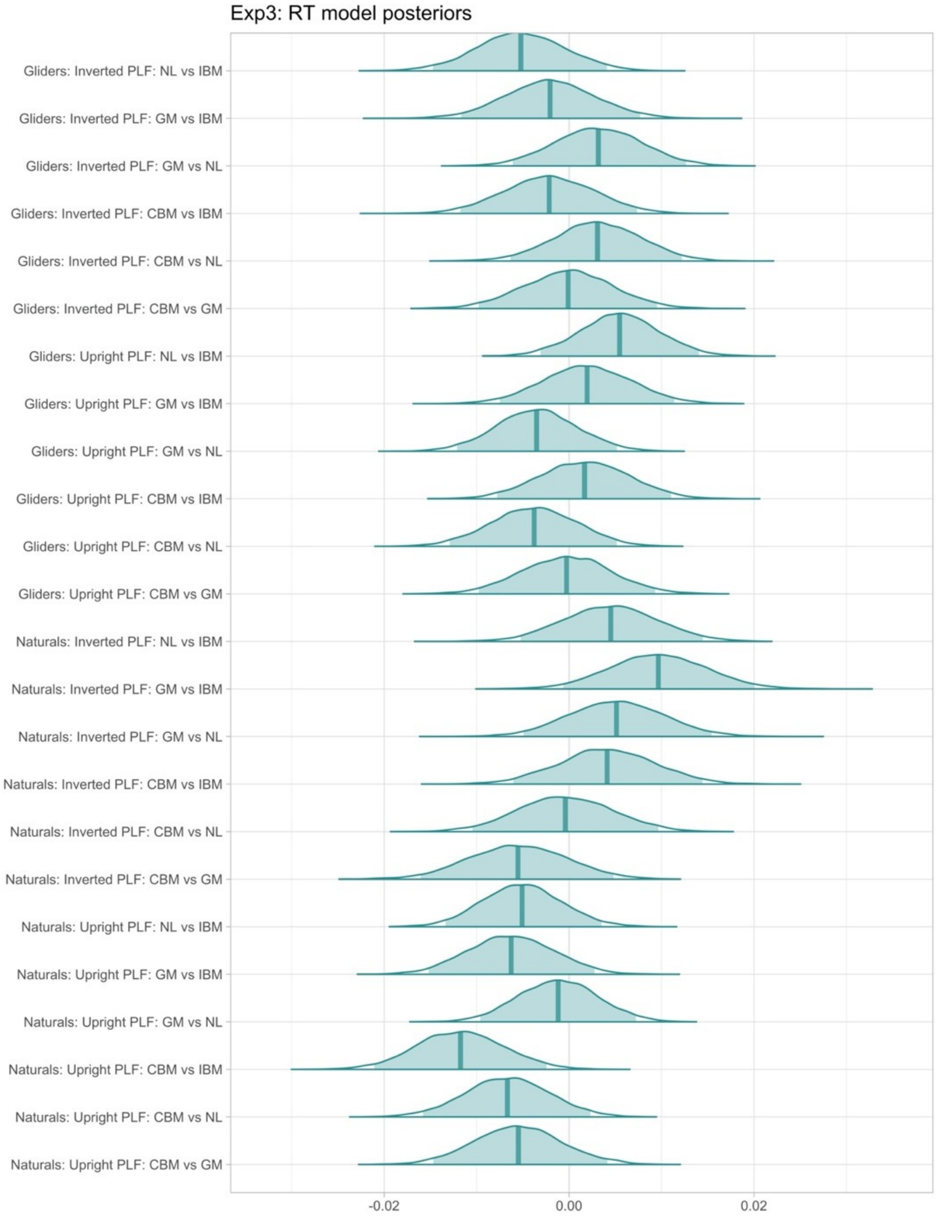


Figure S3.6. Posterior densities for the RT model for Experiment 3. Areas filled in green mark 95% credible intervals. Vertical mark is the mean of the distribution. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, GM – general motion, NL – no language.

CHAPTER 4

The Effect of Language on Biological and General Motion Perception

Abstract

Behavioural studies have shown that linguistic labels can change how we perceive visual input. However, only few studies have investigated the neural signatures of such influences with functional magnetic resonance imaging (fMRI) method. In this study, we examine how lexical cues referring to biological and general motion affect the perception of point-light stimuli depicting either biological motion (a coherent or scrambled point-light figure comprising a person performing an action) or general motion (a random dot motion stimulus with predominantly upward or downward motion vector). Behaviourally, we find that cues congruent with the target facilitate target detection. Looking specifically at the target-relevant motion perception regions, as well as the early visual cortex, derived from the functional localiser, we show that lexical cues can modify activation in both biological and general motion regions. Additionally, we show that in the case of general motion perception, the effect of cues can be seen not only in the motion perception MT/V5 region, but also in the early visual cortex. While the cueing effect is left lateralised for biological motion, it is present in both hemispheres for general motion perception.

In preparation as: Slivac K., Hervais-Adelman A., Flecken M., van den Heuvel M., Hagoort P.(in prep). The Effects of Language on Biological and General Motion Perception.

4.1. Introduction

The perceptual system in the brain is anything but objective and bias-free. Rather, it is susceptible to a number of top-down influences, such as expectations, attention, prior beliefs, etc. (Gilbert & Li, 2013). A particularly notable effect on perception is the one imposed by language. Language plays a role in how we conceptualise and categorise our thoughts and observations, and has been theorised to have the ability to activate a visual representation of what has been linguistically expressed (Barsalou et al., 2003; Binder et al., 2009; Huth et al., 2016).

The most reliable proof of linguistic efficiency in modulating perception comes from a series of behavioural cueing studies. Cueing studies present a well-controlled way of investigating language-perception interaction. By presenting people with linguistic cues followed by a visual target, they measure how the perception of that target changes as a function of the (properties of the) preceding cue. These studies have repeatedly shown that language, when congruent with the visual target, has the ability to increase the accuracy and speed of target detection (Forder & Lupyan, 2019; Francken, Kok, et al., 2015; Lupyan & Thompson-Schill, 2012; Meteyard et al., 2007). It can even make people see targets designed to be invisible to the human eye by means of manipulations such as very short presentation or various masking techniques (Lupyan & Spivey, 2010a; Lupyan & Ward, 2013; Ostarek & Huettig, 2017).

A theory accompanying these findings – the label-feedback hypothesis put forward by Lupyan (2012) – posits that these lexically induced modulatory effects on perception happen due to the ability of linguistic labels to efficiently extract and co-activate a visual template of the features characteristic of the labelled concept (for more detail, see Lupyan, 2012; Lupyan et al., 2020). By doing that, lexical cues are hypothesized to already start activating visual areas relevant for the encoding of visual features of the cue, thus pre-empting the process of target encoding in the case of a match or derailing its perception in the case of a mismatch.

However, while behavioural studies show consistent results confirming the potency of linguistic influence on perception, there is still a lot of debate about the loci of neural activations underlying this influence. Some researchers argue that behaviourally observed effects are mirroring neural patterns of activation, where linguistically delivered concepts cascade into early visual cortex and modulate neural activity therein (Çukur et al., 2013; Lupyan, 2015). And indeed, few neuroimaging studies using techniques such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) have found that language can modulate neural activity in early visual areas (Boutonnet & Lupyan, 2015; Çukur et al., 2013; Noorman et al., 2018; Puri et al., 2009).

Conversely, other researchers maintain that perception is a system unaffected by higher level processes such as language, claiming that any behaviourally observed linguistic effects on perception can and should be ascribed to task-induced bias or decision processes (Firestone & Scholl, 2016; Pylyshyn, 1999). For example, in a well-controlled study, Francken, Kok, et al. (2015) investigated the effect of language on motion perception. Despite showing a behavioural effect of language on the perception of motion direction, they found the neural locus of language-perception interaction to be restricted to the middle temporal gyrus (MTG), traditionally tied to semantic and conceptual processing (Francken, Kok, et al., 2015).

Motion perception is a very interesting case to study in this context. It is a crucial part of our visual system, and can be divided into two distinct subtypes: general motion (also sometimes referred to as abstract or rigid motion) and biological motion perception. While general motion perception is usually associated with simple motion patterns and non-living beings, biological motion perception refers to a more complex type of motion associated with the fluid, self-propagating movement of living beings, such as animals or humans. These two types of motion perception are encoded differently in the brain: while general motion tends to elicit activation in the MT/V5 area, the processing of biological motion perception is linked to the posterior part of the superior temporal sulcus (pSTS) and gyrus (pSTG), the lateral occipito-temporal cortex (LOTC), the fusiform gyrus (FG; including the fusiform body area, FBA), as well as the inferior (IPL) and superior (SPL) parietal lobules (Grosbras et al., 2012; Grossman et al., 2000; Grossman & Blake, 2002; Saygin, 2007).

Importantly, despite several behavioural studies showing the effect of language on the speed and accuracy with which people detect both general and biological motion (Bidet-Ildei et al., 2011; Francken, Kok, et al., 2015; Meteyard et al., 2007; Pavan et al., 2013; Slivac et al., 2021), the neural patterns of activation underlying such effects have not been mapped out yet.

A common way to study motion perception in a controlled manner, without the confound of an explicit shape outline, is to look at point-light kinematograms (PLKs). PLKs are a type of visual stimuli which contain moving dots that can be combined into a recognizable percept, based on their spatial configuration and kinematics. One type of those stimuli is random dot motion (RDM), an aperture full of dots moving in a certain manner or direction, which can be used to study general motion patterns, such as directionality of movement. Another type of point-light stimuli, called point-light figures (PLFs) can be used to study biological motion perception (Johansson, 1973). PLFs consist of 11-13 dots representing key human joints, the location and kinematics of which need to me integrated in space and time in order to obtain a percept of a human figure in action (e.g., walking, dancing, etc.).

The advantage of these stimuli lies in the fact that they are compositional - the final percept that they evoke depends solely on the integration of the patterns of motion of their constituent parts, which are distributed in space without explicit connections or shape outlines. In other words, their perception is based on encoding and combining motion patterns only, without explicitly presented form. This makes them suitable for investigating how our brains process motion neurally, without the confounds of other visual features, such as shape. This also makes them easy to manipulate in a number of different ways - e.g., scrambling the beginning position of the constituent dots, even while preserving their individual kinematics, makes the final percept unidentifiable.

In this study, we test how motion words affect the perception of biological and general point-light stimuli: PLFs and RDMs. We examine how this influence is manifested both behaviourally and neurally, while accounting for the influence of the task itself.

Present Study

In this study, we set out to investigate the neural signatures underlying linguistic cueing effects on the perception of two types of motion – biological and general. In particular, we employed motion language and visual stimuli in order to investigate if and to what extent lexical cues can modify perceptual processes, and what those activation patterns look like under a very specific set of lexical and visual conditions (described in detail below).

To that end, we set up two parts of the experiment, a PLF-task and an RDM-task, aimed at activating two sets of regions within visual cortex – biological motion perception and general motion perception, respectively. In the PLF-task experiment, we examined to what extent lexical cues affect neural activity in biological motion perception regions. In the RDM-task, we examine to what extent those same lexical cues affect the activity in general motion perception regions.

In both the PLF-task and RDM-task, we presented participants with lexical cues: nouns denoting either general (e.g., *snow*, *smoke*, etc.) or biological (e.g., *walker*, *dancer*, etc.) motion. They were followed by a visual target comprised of a biological motion stimulus – a PLF engaged in an action such as walking or dancing, embedded into a general motion stimulus – an RDM aperture with predominantly upward or downward moving dots. We then asked the participants to sequentially perform the two tasks (PLF-task and RDM-task), while keeping both cues and targets exactly the same across two tasks. The PLF task drew their attention (i.e., biased them) towards biological motion perception, asking them to indicate if the PLF in the middle of the RDM aperture represents a coherent or incoherent figure in motion. The RDM task biased them towards general motion perception, asking them to indicate if the RDM dots around the PLF stimulus are moving predominantly in an upward or downward direction. This manipulation allowed us to look at the effect of lexical cues on both biological and general motion, using the same stimuli. Additionally, we conducted a functional localiser session void of a task, in order to localise regions involved in the processing of point-light biological and general motion stimuli in the brain in an uncued manner (see Methods for details).

Looking at the effect of congruence for each of the two types of motion, biological and general motion respectively, we hypothesize that target-congruent lexical cues will have the strongest effect on biological and general motion perception areas respectively, compared to their incongruent or target-unrelated counterparts. Looking at the effect of bias across two motion cue categories, we expect to see that congruent biological motion cues will increase the activation in biological motion perception regions, whereas general motion cues will increase the activation in the general motion perception regions. Alternatively, it can happen that lexical cues affect only the activation in higher-level, semantic and conceptual areas, with motion perception regions staying unaffected by their presence.

4.2. Methods

Participants

Thirty native speakers of Dutch (18 female; mean age: 23.23) recruited from the Radboud University research participation system (Sona) took part in the experiment. All participants were right-handed and had normal or corrected-to-normal vision and had no reading issues. All the participants gave their informed consent and received financial compensation for their participation. The study followed institutional guidelines of the local ethics committee (CMO region Arnhem-Nijmegen, The Netherlands).

Procedure

The experiment had two sessions, performed on two separate days, within 10 days from each other: thresholding and fMRI session. The participants were given both spoken and written instructions before each of the sessions.

Thresholding Session

The thresholding session was conducted in a behavioural lab, with participants sitting in a dark room in front of a computer, approximately 60cm away from the monitor. Stimuli were displayed on a BenQ monitor (24", 1920 x 1080 resolution, 60Hz refresh rate). The session consisted of three parts: familiarization, PLF thresholding, and RDM thresholding procedures.

During familiarization, all visual targets and lexical cues were presented to the participants in isolation, with the instruction to carefully observe the stimuli. The presentation of the PLFs was accompanied by a one sentence description of the type of action they engaged in, e.g., “Je ziet zometeen een figuur, die wandelt” (“*You will see a walking figure*”).

The order of the PLF and RDM thresholding procedures was counterbalanced across participants. Each of the two thresholding procedures started with two practice blocks, and was followed by 3-4 thresholding blocks (192 trials per block). On each trial, a string (####) cue was presented at the beginning of every trial for 100ms, followed by the motion target presented for 700ms. The participants were given 1.5sec to respond: 700ms (during the visual target presentation) + additional 800ms (fixation cross).

If participants reached all the thresholds (see below for the thresholding procedure description) within three thresholding blocks, the procedure was stopped, and the thresholds were extracted after the third block. If the participant failed to reach one of the 6 thresholds on either of the two thresholding procedures after three blocks, an additional 4th block was introduced for that procedure in order to give the participant a final chance to meet the required thresholds. Trial presentation order was fully randomised.

fMRI Session

The fMRI session was conducted in the MRI lab, with participants lying in the scanner in a dark room. The stimuli were delivered on a BOLDScreen (32'', 1920 x 1080 resolution, 120Hz refresh rate). The fMRI session consisted of three parts: PLF-task cueing experiment, RDM-task cueing experiment, and functional localiser.

The PLF- and the RDM-task cueing experiments were counterbalanced across participants. For each of the two tasks, the experiment started with a short practice block (48 trials, same trial structure as the thresholding session) reminding the participants of what the task and experiment looked like and giving them time to adjust to the scanner setup. After that, 3 experimental blocks were presented (260 trials each, 160 stimulus trials and 100 null 'jitter' trials). Each stimulus trial lasted 2 seconds in total, while null trials lasted anywhere between 1 and 9 seconds. Their presentation order was randomised using optseq2, a tool for automatically scheduling events for rapid-presentation event-related experiments (Greve, 2002). Both the task and the trial structure of the PLF- and RDM-task cueing experiments in the MRI scanner were the same as for their respective thresholding procedures. However, in the scanner, they were presented with either a lexical or non-lexical cue ('lucidatypewriter', font size 18) in the middle of the screen, prior to the visual target, instead of only the string #### that preceded the visual target during the thresholding procedures.

The functional localiser consisted of 8 runs (108 trials each, 50 stimuli trials, 51 null 'jitter' trials, 7 catch trials), during which participants were instructed to pay attention to the stimuli presented on the screen and press a button every time a fixation cross in the middle of the screen turned into an X sign (catch trials). Each stimulus trial lasted for 3sec, starting with a stimulus presentation of 700ms followed by a fixation cross for the rest of the trial. Null trials again lasted anywhere between 1 and 9 seconds, and the stimuli order was randomised using optseq2.

The localiser session was always done at the end of the experiment, after the two cueing experiments. The anatomical scan was taken before the localiser part of the MRI experiment.

Stimuli

All stimuli were generated using the Psychophysics Toolbox(Brainard, 1997) within MATLAB R2016a (MathWorks, Natick, MA). Both lexical cues and visual targets were presented in white (luminance: 160 cd/m^2) on a grey background (luminance: 37 cd/m^2).

Task (PLF and RDM)

Each trial, for both the thresholding procedure and the cueing experiment, consisted of a cue (non-lexical cue only for the thresholding procedures, and either a noun or non-lexical cue for the cueing experiments) followed by a visual target (a PLF embedded into an RDM aperture, see **Figure 4.1.**).

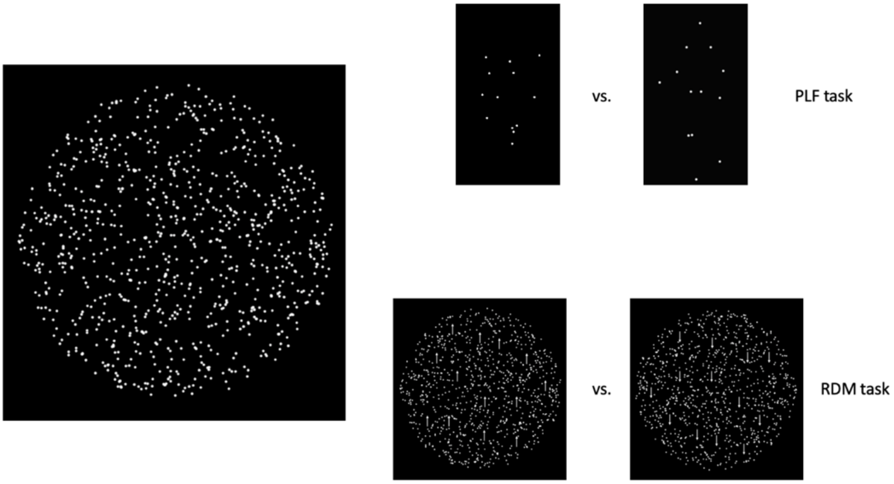


Figure 4.1. Visual stimuli. The visual stimulus was a PLF embedded into the RDM aperture. In the PLF task, participants were asked to pay attention to the figure in the middle of the circle, and indicate whether they detect coherent or incoherent biological motion. In the RDM task, participants were asked to focus on the RDM dots around the figure and indicate whether they are moving in a predominantly upward or downward direction.

For both the PLF-task and the RDM-task cueing experiments, lexical cues were presented in Dutch and consisted of biological motion and general motion cues, as well as one control, no language cue condition. Biological motion cues were either congruent or incongruent with the PLF, while the general motion cues were either directionally congruent or incongruent with the RDM. This resulted in 5 cueing categories: congruent biological motion, incongruent biological motion, congruent general motion, incongruent general motion, and no language cues. For the list of biological motion and general motion cues, see **Table 4.1**. The control, no language cue consisted of the string '####'.

Table 4.1.
Lexical cues

Biological motion	**General motion**
houthakker *(woodcutter)*	hagel *(hail)*
wandelaar *(walker)*	sneeuw *(snow)*
danser *(dancer)*	rook *(smoke)*
	stoom *(steam)*

The cues were presented in Dutch. English translation is provided in the brackets. The control, non-linguistic cue was '####'. Biological motion cues were either congruent or incongruent with the PLF stimulus. General motion cues were either directionally congruent or incongruent with the RDM stimulus.

The visual targets consisted of 13 white dots comprising a point-light figure (PLF; size: 3.59 – 4.36cm (horizontal)*5.95 – 6.57cm (vertical); speed: 30 frames/second) embedded in a random dot motion mask (RDM), with circular aperture (22*23.5cm; number of dots: 866; dot size: 0.528*0.528mm, dot motion speed: 0.528mm/frame, dot lifetime: 10 frames, at 30 frames/second), presented in the middle of the screen.

Three PLF types, performing three types of actions: wood-cutting (with an axe), walking, and dancing, were selected from an action database (Vanrie & Verfaillie, 2004). On every trial, PLFs were presented, facing to the left or to the right in sagittal view (45°), in the centre of the RDM aperture. They were shown either in their coherent form, comprising a human figure in action or in a scrambled form, where the initial locations of the landmark dots were randomly offset within the perimeter of the coherent PLF, while their individual kinematics were preserved. This manipulation rendered the target unrecognizable as a coherent human figure in motion.

The RDM dots were moving either predominantly in an upward or downward direction – on every trial, a certain proportion of dots moved coherently in one of the two directions while the rest of the incoherent dots were re-drawn in a random location at every monitor refresh. The percentage of the coherently moving dots in either direction was determined by the thresholding procedure (see below).

Individual thresholding levels (i.e., the percentage of RDM dots surrounding the PLF, moving coherently in an upward or downward direction) were determined for the PLF-task and the RDM-task cueing experiments separately. Both had the exact same visual stimuli (a PLF embedded in an RDM), but the difference was in the task set for participants.

In the PLF-task thresholding procedure and cueing experiment, participants were asked to indicate with a button press whether the PLF target is coherent or scrambled (i.e., the PLF was the target while the RDM was the distractor/mask). In the RDM-task thresholding procedure and cueing experiment, they were asked to indicate with a button press whether the RDM target was moving coherently upwards or downwards (the RDM was the target while the PLF was the distractor).

The PLF thresholding procedure was done for each of the three PLF types and each of the two RDM directions per participant using a Bayesian adaptive staircase procedure (QUEST; Watson & Pelli, 1983). For every action, the threshold was collapsed across left and right PLF orientation, and across coherent and scrambled figures. At the end of the staircase procedure, we extracted six thresholds for every participant which reflected the masking level at which the three actions embedded in either upward or downward moving RDM yielded approximately 75% accuracy on a biological motion discrimination task.

The RDM thresholding procedure was done separately for each distractor – three PLF types, coherent and scrambled, and collapsed across upward and downward RDM direction. We used the same QUEST procedure as for the PLF task. At the end of the staircase procedure, we extracted six thresholds for every participant which reflected the RDM coherence level at which participants performed 75% accurately on a general motion discrimination task, given the six PLF distractors.

Participants who did not reach the 75% accuracy performance level on all twelve thresholding conditions described above, even when all the noise dots were moving coherently (i.e., at the easiest level of performance) during the thresholding procedure were excluded from the experiment.

Localiser

For the fMRI localiser session, both lexical and visual stimuli presented throughout the cueing experiment were shown in isolation, in the centre of the screen. They included all motion words used as cues in the cueing experiment (biological motion and general motion nouns), the PLFs of the three actions selected for the cueing experiments, in their coherent and scrambled form, the RDMs moving in an upward and downward direction, as well as a completely incoherent RDM (all dots moving randomly) and a static RDM (an image of an RDM without motion). The localiser session was performed in order to obtain areas in the brain responsible for encoding stimuli presented throughout the experiment in isolation, without being influenced by the task, distractors or masks, and cues.

fMRI Data Acquisition

Anatomical and functional images were acquired on a 3T Prisma scanner (Siemens, Erlangen, Germany), using a 32-channel head coil. Anatomical images were acquired using a sagittal 3D sparse MP2RAGE sequence with the following parameters: TR/TI1/TI2 = 5000/700/2500ms, TE = 3.1ms, echo spacing = 7.6ms, flip angles (a1/a2) = 4°/5°, iPAT = 2, voxel size = 0.8x0.8x0.8. Functional images were acquired using a whole-brain T2*-weighted multiband-6 sequence with the following parameters: TR/TE=1000/34ms, slices = 66, voxel size: 2.0x2.0x2.0, flip angle = 60°, phase encoding direction = A>>P, FOV = 210mm, BW = 2090 Hz/Px.

4.2.1. Data Analysis

Behavioural Analysis

PLF Task

Behavioural data analysis was performed on 28 participants, out of 30 collected. Two participants performed at chance level on the PLF task in the scanner (54.7% accuracy and 54.2% accuracy respectively, when 55.4% accuracy is considered above chance for our study), despite performing at the 75% accuracy level during the thresholding session, and were therefore excluded from the analysis. Before the analyses, trials with reaction times (RTs) 2.5 SD or more from the grand mean were excluded (trials with RTs above 1652.371ms and below 761.5375ms). This resulted in the exclusion of 148 out of 14400 trials (1% of trials).

For accuracy and RT measures, we were interested in how the lexical cue categories (congruent and incongruent biological motion, and congruent and incongruent general motion) influenced both the detection of coherent and rejection of scrambled PLFs. We expected the cues that exert a facilitatory effect on the detection of a coherent target to also be detrimental to the rejection of scrambled targets (i.e., would lead to a higher false alarm rate) and vice versa. In order to test that, the analysis measured the effect of congruent biological motion cues compared to other 3 lexical cues and the (control) no language cue, nested under two PLF coherence conditions: coherent and scrambled.

Both accuracy (binomial distribution) and RT (Gaussian distribution, with log-transformed RTs) mixed effect models as implemented in the lme4 package in R were fitted with the hypothesis-driven, non-singular structure supported by our data (Bates et al., 2015). The final model consisted of the predictor ‘cue category’ (5 levels: congruent biological motion, incongruent biological motion, congruent general motion, incongruent general motion, and no language) nested under the predictor ‘PLF coherence’ (2 levels: coherent, scrambled) as fixed effects, and by-subject and by-item random intercepts and slopes for PLF coherence as random effects. For fixed effects, simple effects coding was used, with the PLF coherence predictor coded: coherent as 0.5, scrambled as -0.5, and the cue category predictor coded: congruent biological motion cues were base coded as -0.2, contrasting condition of each column as 0.8.

RDM Task

Data analysis was performed on 30 completed data sets. Trials with reaction times (RTs) 2.5 SD or more from the grand mean were excluded (trials with RTs above 1639.766ms and below 555.7709ms). This resulted in the exclusion of 181 out of 14400 trials (1.2% of trials).

For the same reason as in the PLF analysis, we measured the effect of congruent general motion cues compared to other 3 lexical cues and the (control) no language cue, this time nested under two RDM directions: upward and downward motion.

Both accuracy (binomial distribution) and RT (Gaussian distribution, with log-transformed RTs) mixed effect models were again fitted with the hypothesis-driven, non-singular structure supported by our data. The final model consisted of the predictor ‘cue category’ (5 levels: congruent biological motion, incongruent biological motion, congruent general motion, incongruent general motion, and no language) nested under the predictor ‘RDM direction (2 levels: upward, downward) as fixed effects, and by-subject and by-item random intercepts and slopes for RDM direction as random effects. For fixed effects, simple effects coding was used, with the RDM direction predictor coded: upward as 0.5, downward as -0.5, and the cue category predictor coded: congruent general motion cues were base coded as -0.2, contrasting condition of each column as 0.8.

fMRI Analysis

fMRI Data Preprocessing

Data were preprocessed using fMRIPrep 20.2.1 (Esteban et al., 2019), which is based on Nipype 1.5.1 (Gorgolewski et al., 2011).

Anatomical Data

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU), skull-stripped and brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w with fast (FSL 5.0.9, Zhang et al., 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 6.0.1; Dale et al., 1999). Volume-based spatial normalization to a standard space (MNI152NLin2009cAsym) was performed using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov et al., 2009).

Functional Data

For each of the 14 BOLD runs per subject (PLF task, RDM task, localiser), the following preprocessing was performed: First, a reference volume and its skull-stripped version were generated by aligning and averaging 1 single-band references (SBRefs). A B0-nonuniformity map (or fieldmap) was estimated based on a phase-difference map calculated with a dual-echo GRE (gradient-recall echo) sequence, co-registered to the target EPI (echo-planar imaging) reference run and converted to a displacements field map. Based on the estimated susceptibility distortion, a corrected EPI (echo-planar imaging) reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using mcflirt (FSL 5.0.9, Jenkinson et al., 2002). BOLD runs were slice-time corrected and resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. The BOLD time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. 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. This mask is obtained by dilating a GM mask extracted from the FreeSurfer\u2019s aseg segmentation and resampled into BOLD space and binarized by thresholding at 0.99. Frames that exceeded a threshold of 0.5mm FD or 1.5 standardised DVARS were annotated as motion outliers.

Localiser Analysis

In order to obtain regions of interest for the ROI task analyses, mass univariate analysis was performed on the localiser data, using FSL with the Nipype interface as implemented in Python (Gorgolewski et al., 2011).

The functional images were smoothed with SUSAN, using a Gaussian kernel (full-width at half-maximum of 6mm). At first level, 17 stimuli regressors were included in the analysis, corresponding to the 17 stimuli types presented to the participants during the localiser session. Beside the stimuli regressors, the design matrix also included the following confounding regressors calculated via fMRIPrep: framewise displacement, 6 standard motion parameters, 6 anatomical CompCor components, and 2 cosines, as well as motion outliers.

FSL FEAT was used to fit voxel-wise general linear models (GLM) to each participant's run data in an event-related approach. For the model, events were set as instantaneous (duration = 0sec) and convolved with double gamma haemodynamic response function. FSL's fixed effects (FLAME's fixed effects) analysis was used to combine data across runs. Finally, across participants, data were combined using FSL's mixed effects analysis (FLAME 1). Gaussian random-field cluster thresholding was used to correct for multiple comparisons, using the updated default settings of FSL 5.0.11, with a cluster formation threshold of $p<0.001$ (one-sided; that is, $z = 3.1$) and cluster significance threshold of $p < 0.05$.

ROI Analysis

The ROI analysis was performed in order to investigate the effects of lexical cues in motion-sensitive areas of the brain. The analysis was done for both the PLF task and RDM task, respectively. The biological motion and general motion ROIs were obtained empirically from the functional localiser session, and included higher-level motion perception regions (biological motion regions for the PLF task and general motion regions for the RDM task), as well as lower-level visual regions for both tasks.

PLF Task

In order to get a pattern of activation in response to biological motion only, we subtracted the activation pattern of the moving RDM stimuli from that of the PLF stimuli in the functional localiser. The resulting two ROIs were the left and right lateral occipito-temporal cortex (LOTC), the inferior parietal lobule (IPL) and parts of the superior temporal lobule (SPL, in the right hemisphere), known to respond to biological stimuli. The early visual ROI encompassed the most restrictive region activated by all point-light stimuli during the localiser session, encompassing the bilateral calcarine sulci, the posterior lingual gyri, and the inferior occipital gyri.

For each ROI, mean parameter estimates were extracted from each participants' parameter estimate maps, representing each of the five cueing categories (CBM, IBM, CGM, IGM, NL) for two PLF coherence conditions (coherent and scrambled) compared to baseline. This resulted in ten parameters of interest. The parameter estimates were then divided by 100, to yield the measure approximating signal percent change relative to baseline (Mumford, 2007). For each ROI, these data were then analysed with linear mixed effects models, with the same fixed effects structure and contrasts as the behavioural models for the PLF task, and by-subject random intercepts and slopes for PLF coherence as random effects.

RDM Task

In order to get a pattern of activation in response to general motion only in the localiser session, we subtracted the activation pattern of the static RDM stimulus from that of the moving RDM stimuli. From there we extracted two ROIs – the left and right MT/V5 region, known to respond to general motion stimuli, such as random dot motion kinematograms. Additionally, the same early visual cortex ROI as in the PLF task was included in the RDM task analysis.

The parameter estimates were extracted for each ROI the same way as for the PLF task, and the linear mixed effects model had the same fixed effects formula and contrasts as the behavioural models for the RDM task, and by-subject random intercepts and slopes for PLF coherence and random effects, and by-subject random intercepts. By subject random slopes for RDM direction were removed from the model's random effects to avoid singular fits.

Exploratory Whole-Brain Analysis

In addition to the ROI analyses, we also performed an exploratory whole-brain mass univariate analysis, in order to get a complete profile of activation patterns subserving language-perception interaction.

PLF Task

For the PLF task data, we performed the same type of mass univariate analysis as for the functional localiser. Smoothing and nuisance regressors were the same as for the localiser data, while the regressors of interest were chosen with respect to our contrasts of interest (see below).

We were interested in the effects of our lexical cues, hypothesized to bias the perceptual activation towards the PLF target, compared to no language cues and those lexical cues hypothesized to bias perceptual activation away from the target, on the perception of coherent and scrambled PLF targets respectively. For that reason, our analysis contrasted congruent biological motion cues (CBM) against (control) no language cues (NL), to test for the effect of biological motion labels compared to no labels. To test for the effect of congruence for biological motion cues biasing perception towards the target, we also looked at the effect of CBM cues compared to incongruent biological motion cues (IBM). Finally, to test for the effect of cues biasing perception towards the target compared to cues biasing away from the target, we contrasted the effect of CBM cues to both congruent (CGM) and incongruent (IGM) general motion cues.

RDM Task

For the RDM task, the same mass univariate analysis was performed with the same smoothing and nuisance regressors as for the functional localiser and PLF task.

We were again interested in the effects of our lexical cues, hypothesized to bias the perceptual activation towards the RDM target compared to no language cues and those hypothesized to bias perceptual activation away from the target, on the perception of upward and downward RDM targets respectively. For that reason, our analysis contrasted congruent general motion cues (CGM) against (control) no language cues (NL), incongruent general motion cues (IGM), and two cue categories biasing perception away from the target and towards the distractor – congruent biological motion cues (CBM) and incongruent biological motion cues (IBM).

4.3. Results

4.3.1. PLF Task

Behavioural Results

Accuracy

For coherent PLFs (mean accuracy: 61.19%; see **Figure 4.2.**), the accuracy model showed no significant results. For scrambled PLFs (mean accuracy: 79.73%), the accuracy model revealed a significantly lower accuracy rate for congruent biological motion cues compared to no language cues (CBM vs. NL; estimate = -0.312, SE = 0.102, z = -3.049, p = 0.002) and compared to incongruent biological motion cues (CBM vs. IBM; estimate = -0.303, SE = 0.102, z = -2.960, p = 0.003).

RTs

For coherent PLFs (mean RT: 895.95ms; see **Figure 4.2.**) the model revealed faster RTs for congruent biological motion cues compared to incongruent biological motion cues (CBM vs. IBM; estimate = -0.018, SE = 0.007, z = -2.725, p = 0.006). For scrambled PLFs (mean RT: 920.70ms) the model showed slower RTs for congruent biological motion cues compared to congruent general motion cues (CBM vs. CGM; estimate = 0.014, SE = 0.006, z = 2.350, p = 0.018).

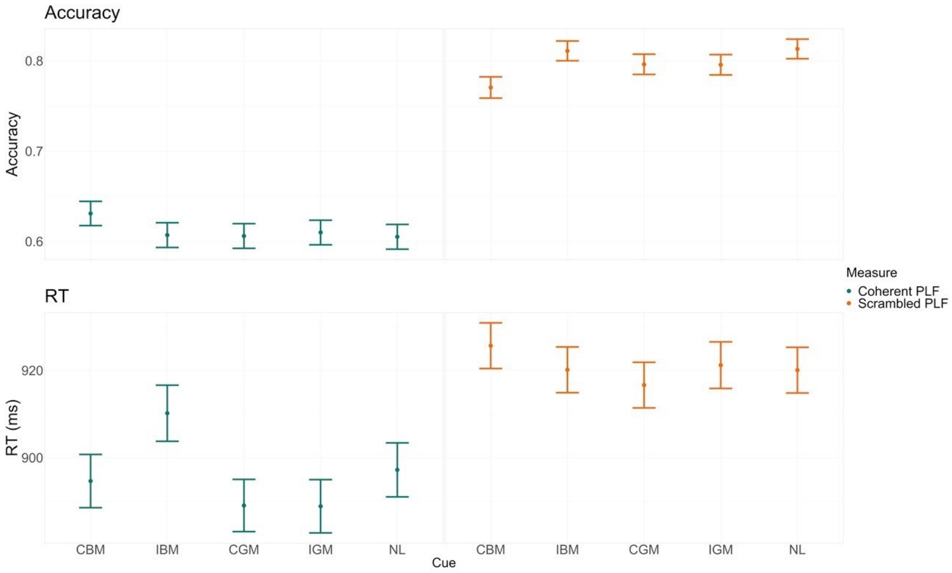


Figure 4.2. Behavioural results for the PLF task. Accuracy and RT. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, CGM – congruent general motion, IGM – incongruent general motion, NL – no language.

To sum up, behavioural results revealed that congruent biological motion cues led to a higher false alarm rate compared to no language cues and incongruent biological motion cues. In other words, under the influence of biological motion cues congruent with the target, participants were more likely to report seeing a coherent PLF in action even when that PLF was made nonsensical by scrambling.

Additionally, we found that congruent biological motion cues sped up the perception of coherent PLFs compared to their incongruent counterparts. However, when the PLF target was scrambled, reaction times were slower for congruent biological motion cues compared to distracting general motion cues, directionally congruent with the RDM mask. With these reaction time results, we again see the dissociation of facilitatory and detrimental effects of lexical cues – a boost in speed when the target matches the label (i.e., depicts the same action), and a lag in correct rejection responses when the target is made nonsensical by means of scrambling.

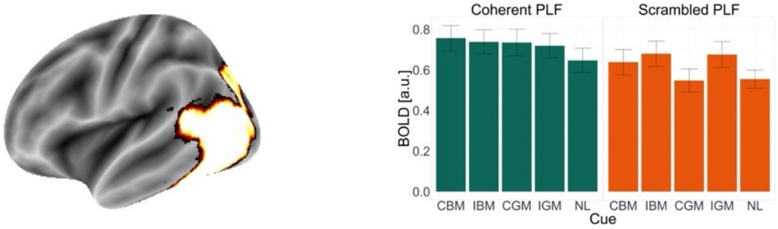
fMRI Results

ROI Analysis Results

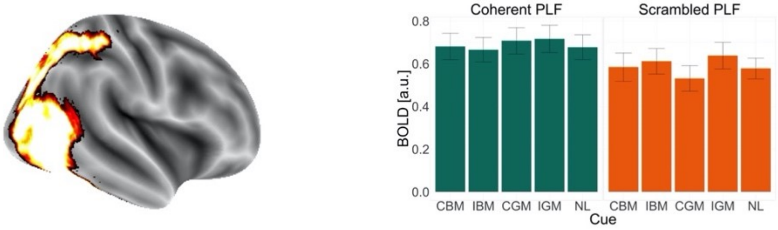
In the left biological motion ROI (see **Figure 4.3.A**), for coherent PLFs, the mixed effects model showed a significant increase in activation for CBM cues compared to NL cues (CBM vs. NL; estimate = 0.11, SE = 0.039, t = 2.804, p = 0.005). For scrambled PLF targets, the model revealed a significant increase in activation for CBM cues compared to NL cues (CBM vs. NL; estimate = 0.084, SE = 0.039, t = 2.138, p = 0.034) and compared to CGM cues (CBM vs. CGM; estimate = 0.091, SE = 0.039, t = 2.312, p = 0.022).

In the right biological motion ROI (see **Figure 4.3.B**), the mixed effects model showed no significant changes in activation as a function of the cueing conditions. In the early visual cortex (see **Figure 4.3.C**), no significant changes in activation were found as a function of the cueing conditions.

A) Left ROI (Biological Motion)



B) Right ROI (Biological Motion)



C) Visual ROI (Biological Motion)

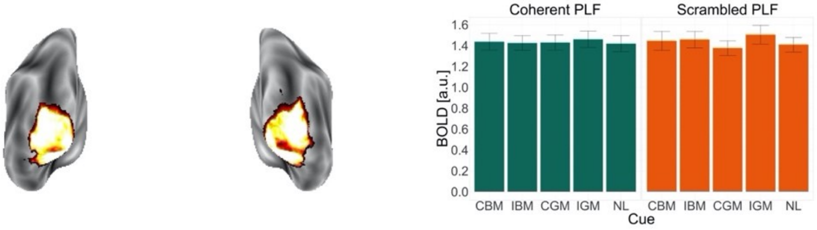


Figure 4.3. Left and right biological motion, and early visual cortex ROI. The biological motion regions activated were left (**A**) and right (**B**) hemisphere lateral occipito-temporal cortex (LOTC), the inferior parietal lobule (IPL), and the superior parietal lobule (SPL; in the right hemisphere). The early visual region (**C**) involved the bilateral calcarine sulci, the posterior lingual gyri, and the inferior occipital gyri. The bar plots show parameter estimates per cue category and PLF coherence, extracted from ROIs.

To sum up, we found that congruent biological motion cues lead to an increase in activation in the left biological motion ROI, but not right, or the early visual cortex. Specifically, congruent biological motion cues increased the activation in the left ROI compared to no language cues for both coherent and scrambled PLF targets, and compared to (distracting) congruent general motion cues for scrambled PLF targets. These findings show that linguistic cues can modulate neural activity in biological motion perceptual regions, but only in the left-hemisphere. The left-hemisphere only effects confirm the previously noticed tendency for the lexical top-down influence to be pronounced more locally, within the hemisphere dominant for language processing (Francken, Kok, et al., 2015).

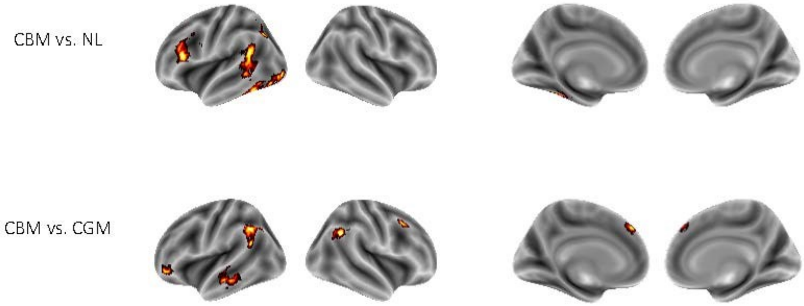
Whole-Brain Analysis Results

The whole-brain univariate analysis revealed the following patterns of activation (see Supplementary **Table S4.1**). On trials with a coherent PLF target (see **Figure 4.4.A**), the CBM vs. NL cue contrast revealed significant clusters with peaks in the left fusiform gyrus (FG), the left posterior middle temporal gyrus (pMTG) and the superior temporal sulcus (pSTS), the left inferior frontal gyrus (IFG), and the left inferior (IPL) and superior (SPL) parietal lobules. The CBM vs. CGM cue contrast for the same type of (coherent PLF) target revealed significant clusters with peaks in the left IPL; the right angular gyrus (AG), the left frontal medial orbital gyrus, the left frontal superior medial gyrus, the left MTG and the right middle frontal gyrus.

On trials with a scrambled PLF target (see **Figure 4.4.B**), the CBM vs. NL contrast revealed clusters with peaks in the left IFG, the right supramarginal gyrus, the left IPL, the left pMTG and pSTS, and the left inferior temporal gyrus (ITG). CBM vs. CGM contrasts for the same target revealed clusters with peaks in the right posterior superior temporal gyrus (pSTG), the left pSTG and pSTS, the right MTG, and the left IPL and SPL.

In sum, our exploratory whole-brain analysis revealed clusters in the frontal, temporal and parietal regions involved in higher level functions such as semantic and conceptual processes, social cognition, and working memory.

A) Coherent PLF



B) Scrambled PLF

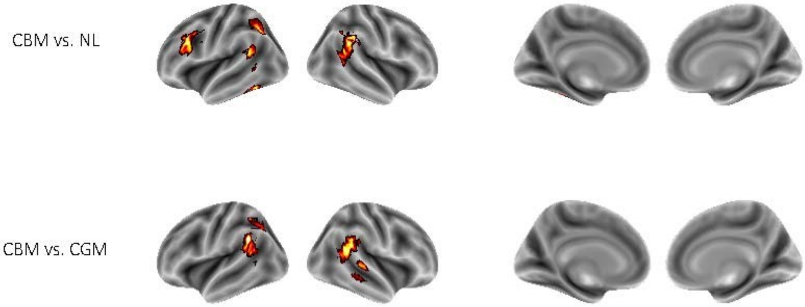


Figure 4.4. Whole-brain clusters (PLF task). Significant clusters presented for coherent (**A**) and scrambled (**B**) PLF targets, for CBM vs. NL and CBM vs. CGM contrasts. Gaussian random-field cluster thresholding was performed, with a cluster formation threshold of $p<0.001$ (one-sided; $z = 3.1$) and cluster significance threshold of $p < 0.05$.

4.3.2. RDM Task

Behavioural Results

Accuracy

For upward RDM (mean accuracy: 68.89%, see **Figure 4.5.**), the accuracy model showed no significant results. For downward RDM (mean accuracy: 73.78%), the accuracy model revealed a significantly lower accuracy rate for congruent general motion cues compared to congruent biological motion cues (CGM vs. CBM; estimate = -0.272, SE = 0.091, z = -3.001, p = 0.003) and incongruent biological motion cues (CGM vs. IBM; estimate = -0.184, SE = 0.09, z = -2.047, p = 0.04).

RTs

For upward RDM (mean RT: 788.55ms, see **Figure 4.5.**), the RT model showed no significant results. For downward RDM (mean RT: 770.27ms), the RT model revealed faster RTs for congruent general motion cues compared to no language cues (CGM vs. NL; estimate = -0.024, SE = 0.008, z = -3.093, p = 0.002), incongruent general motion cues (CGM vs. IGM; estimate = -0.019, SE = 0.008, z = -2.436, p = 0.015), and congruent biological motion cues (CGM vs. CBM; estimate = -0.016, SE = 0.008, z = -2.070, p = 0.038).

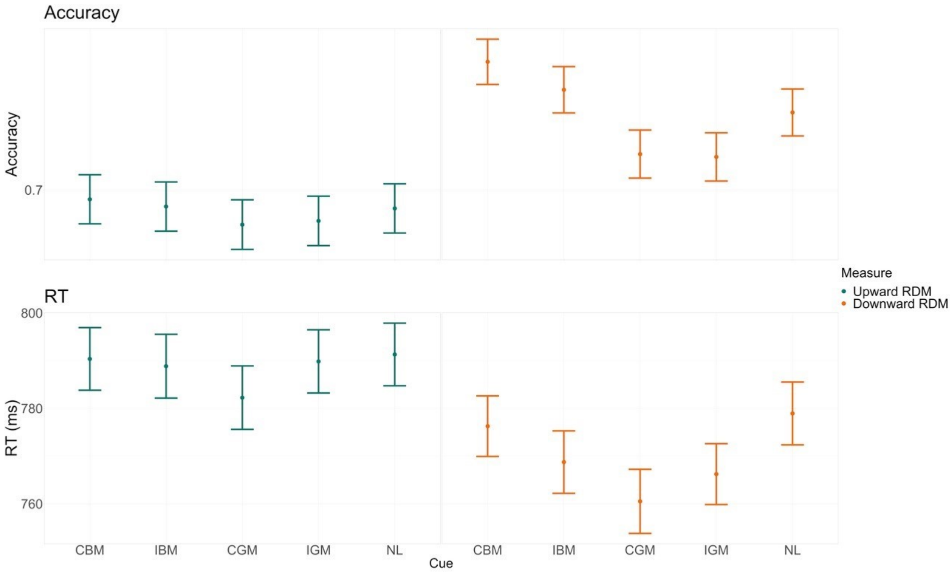


Figure 4.5. Behavioural results for the RDM task. Accuracy and RT. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, CGM – congruent general motion, IGM – incongruent general motion, NL – no language.

In sum, behavioural analysis showed that both distractor cues – congruent and incongruent biological motion cues – led to higher accuracies in detecting downward motion than target-focusing congruent general motion cues. No effect of cues was found for upward moving RDMs. As a reminder, each RDM stimulus with upward or downward moving dots had a distractor PLF hidden in the middle of an aperture. It could be then that these surprising results were observed because the distractor figure provided a reference point against which the directionality of the movement around it could be judged, thus leading to a better judgement of general motion.

Conversely, reaction time results confirmed out hypothesis in showing that general motion cues directionally congruent with the RDM target led to faster responses compared to control no language cues, incongruent general motion cues and (distracting) congruent biological motion cues. This effect was again observed only for downward motion, but not upward. The implications of those findings are discussed in the Discussion below.

fMRI Results

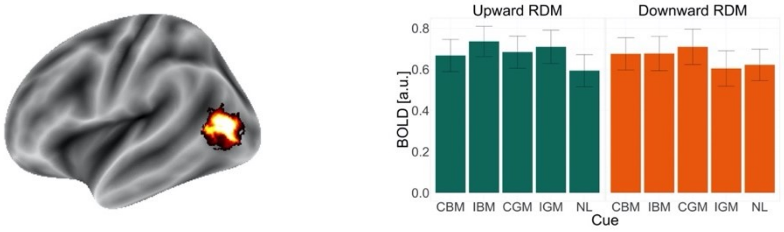
ROI Analysis Results

In the left general motion ROI (see **Figure 4.6.A**), for upward RDM targets, the mixed effects model showed a marginally significant increase in activation for CGM cues compared to NL cues (CGM vs. NL; estimate = 0.09, SE = 0.047, t = 1.934, p = 0.054). For downward RDM target, the model revealed a significant increase in activation for CGM compared to IGM cues (CGM vs. IGM; estimate = 0.105, SE = 0.047, t = 2.249, p = 0.02).

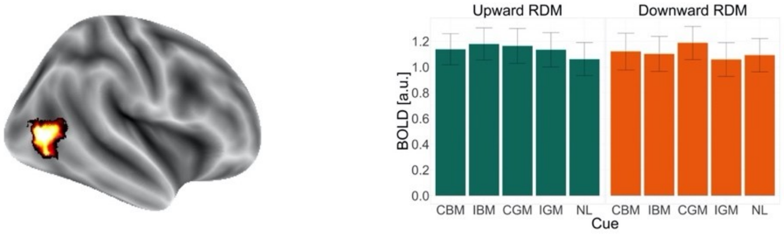
In the right general motion ROI (see **Figure 4.6.B**), for upward RDM targets, the mixed effects model again showed a significant increase in activation for CGM cues compared to NL cues (CGM vs. NL; estimate = 0.103 SE = 0.049, t = 2.104, p = 0.036). For downward RDM, the model revealed marginally significant increase in activation for CGM cues contrasted with NL cues (CGM vs. NL; estimate = 0.096, SE = 0.049, t = 1.957, p = 0.051), and with IGM cues (CGM vs. IGM; estimate = 0.129, SE = 0.049, t = 2.644, p = 0.009).

In the early visual cortex (see **Figure 4.6.C**), for upward RDM targets, the model revealed a significant increase in activation for CGM cues compared to NL cues (CGM vs. NL; estimate = 0.125, SE = 0.045, t = 2.763, p = 0.006), and compared to CBM cues (CGM vs. CBM; estimate = 0.095, SE = 0.045, t = 2.088, p = 0.038). No significant results were found for downward RDM targets.

A) Left ROI (General Motion)



B) Right ROI (General Motion)



C) Visual ROI (General Motion)

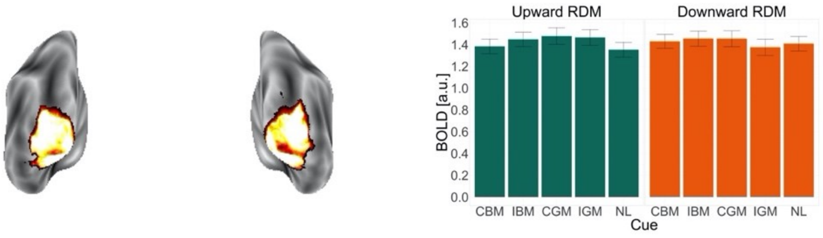


Figure 4.6. Left and right ROI, and the early visual cortex ROI. The regions activated by the RDMs in the functional localiser session were left (**A**) and right (**B**) posterior middle temporal gyrus (MT/V5). The early visual region (**C**) involved bilateral calcarine sulci, the posterior lingual gyri, and the inferior occipital gyri. The bar plots show parameter estimates per cue category and PLF coherence. The bar plots show parameter estimates per cue category and RDM direction, extracted from the ROIs.

In sum, the region of interest analysis found that congruent general motion cues lead to an increase in activation in both the left and right MT/V5 region, compared to no language cues and incongruent general motion cues. Interestingly, this effect compared to no language cues was observed only for upward RDMs, while the effect compared to incongruent general motion cues was present only for the downward RDMs.

Whole-Brain Analysis Results

The whole-brain univariate analysis revealed the following patterns of activation (see Supplementary **Table S4.2.**). On trials with an upward RDM target (see **Figure 4.7.**), the CGM vs. NL contrast revealed clusters with peak activations in the IPL and postcentral gyrus, the right cuneus, the left superior occipital gyrus (SOG), the left cerebellum and the FG, the right cerebellum, the left inferior occipital gyrus (IOG), and the right postcentral gyrus. The IGM vs. CGM contrasts revealed clusters in the left and right AG, and the left MTG. The CGM vs. CBM contrast for the same target revealed a cluster with the peak activation in the bilateral postcentral gyri.

On trials with a downward RDM target (see **Figure 4.8.**), the CGM vs. NL contrasts revealed clusters with peaks in the left precentral gyrus. The NL vs. CGM contrast revealed a cluster in the left caudate nucleus. The CGM vs. IGM contrast revealed a cluster with a peak in the bilateral supplementary motor area. The CGM vs. IBM contrasts revealed a cluster with a peak in the right IFG.

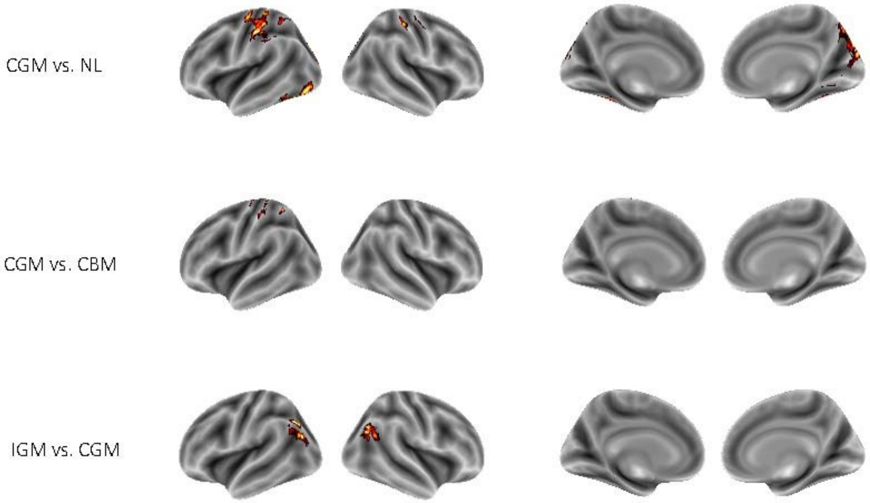


Figure 4.7. Whole-brain analysis from the RDM task (upward RDM). Significant clusters were obtained from gaussian random-field cluster thresholding, with a cluster formation threshold of $p < 0.001$ (one-sided; $z = 3.1$) and cluster significance threshold of $p < 0.05$.

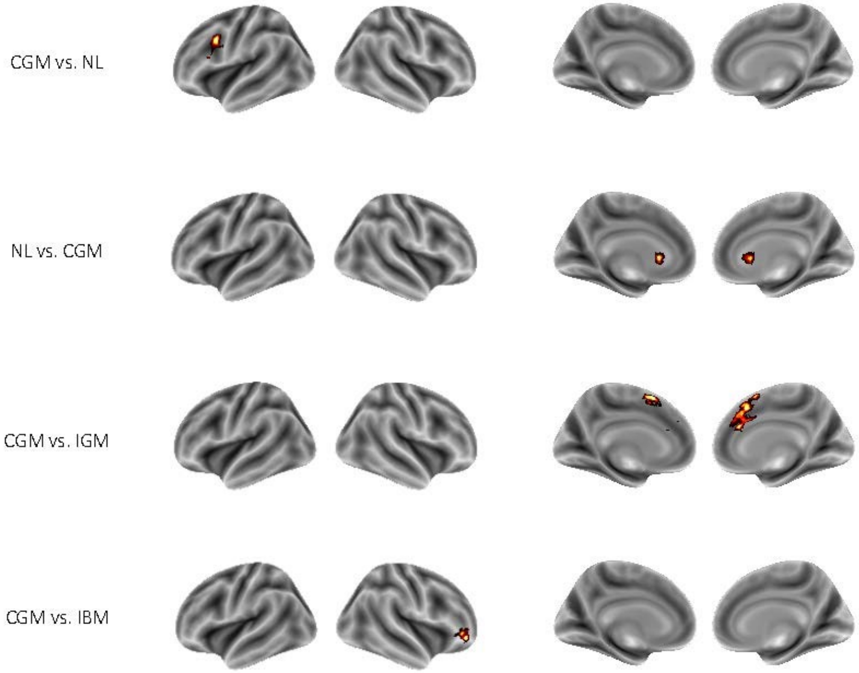


Figure 4.8. Whole-brain analysis from the RDM task (downward RDM). Significant clusters were obtained from gaussian random-field cluster thresholding, with a cluster formation threshold of $p < 0.001$ (one-sided; $z = 3.1$) and cluster significance threshold of $p < 0.05$.

Taken together, the whole-brain analysis resulted in no significant increase in activation in regions traditionally seen as motion perception regions, such as the MT/V5 area. Instead, a more general visuo-motor pattern of activation was observed, alongside the attentional, saccadic and working memory regions, usually involved in conflict resolving and directional decision-making processes.

4.4. Discussion

In this study, we set out to examine the effect of language on two types of motion perception: biological and general motion. Specifically, we looked at how lexical cues congruent with a visual motion target affect the perception of that target on a behavioural and neural level.

Effects of Language on Biological Motion Perception

In the experiment with the PLF task, we looked at the effects of lexical motion cues on biological motion perception. The results showed that biological motion cues semantically congruent with the target lead to poorer discrimination between coherent and scrambled PLFs. Specifically, participants were more likely to report seeing a coherent PLF in action even when that PLF was made nonsensical by scrambling, when cued by coherent biological motion nouns compared to their incongruent counterparts. This finding has been reported before in a biological motion detection study using coherent and scrambled PLF targets (Slivac et al., 2021). Given the nature and manipulation of our visual targets (i.e., scrambled stimuli still had preserved biological motion kinematics), our findings suggest that biological motion labels can intercept with visual configuration processes, and bias people towards erroneously combining point-light dots into a percept congruent with what has been labelled, even in the absence of a coherent target.

Additionally, reaction time results showed a dissociation of facilitatory and detrimental effects of lexical cues congruent with the target – a boost in speed when the target was coherent, and slowing down of correct rejection responses when that figure is made nonsensical by means of scrambling. In the context of our visual targets, the longer reaction times observed with scrambled targets could reflect the extra time needed to overcome the erroneous visual representation evoked by the cues in order to reject the nonsensical target. While studies before reported an increase in discriminability accuracy between two semantically different stimuli (e.g., upward vs. downward motion; Meteyard et al., 2007), no other studies looked at the discriminability within the same semantic category (e.g., dancer) for sensical (coherent dancer) and nonsensical (scrambled dancer) stimuli.

Taken together, our results follow the well documented facilitatory effect of linguistic cues on target perception in the case of cue-target congruence (i.e., overlap in features expressed by the cue semantics and incorporated in the target). However, we also show that, in ambiguous settings where the target is made nonsensical by scrambling, those same cues can mislead participants into constructing wrong percepts. Given the nature of our stimuli, lexical top-down influences seem to affect perceptual stages involved in the integration of point-light dots into a percept (i.e., a figure in motion).

In order to examine whether linguistic labels can modulate neural activation in biological motion and early visual regions, we conducted an ROI analysis in those regions. We found that congruent biological motion cues lead to an increase in activation in the left biological motion ROI compared to no language cues for both coherent and scrambled PLF targets, and compared to mask-congruent general motion cues on trials with scrambled PLF targets. However, no effect of linguistic cues was found in either the right biological motion ROI or the early visual cortex.

As a reminder, the control condition was a string of hash signs rather than a neutral word, suggesting that target-congruent lexical cues may be able to increase activation in biological motion areas compared to cues that are not linguistic in nature. However, when it comes to congruence within lexical cue categories, we only found an increase in the magnitude of activation for scrambled PLF targets. These findings suggests that in the absence of meaningful visual input (that merely hints at an action via preserved kinematics), top-down lexical influences may shift the activation in biological motion regions towards the cued concept. In other words, if the target is not immediately recognisable, the top-down effect of language on perception may become more prominent and the cue-evoked activation bias in biological motion regions may become stronger. In sum, this finding suggested that unclear targets may be more susceptible to top-down lexical influences.

Further, the cue-evoked effect exclusive the left hemisphere suggests that linguistic cues can modify neural activation in earlier biological motion regions, but only in the hemisphere dominant for language processing. The left-hemisphere only effects confirm the previously reported tendency for the lexical top-down influence to be more pronounced locally, within the hemisphere dominant for language (Francken, Kok, et al., 2015). This left-lateralisation can be a reflection of a possibility that linguistic top-down information must travel across corpus callosum to get to the right hemisphere and the attenuated or non-existent effect there is due to that distance. Alternatively, it can indicate that the process of categorisation in general is left-lateralised (Drivonikou et al., 2007).

In this study, we observed an increase in neural activation in motion-specific perceptual regions as a function of lexical cues. An increase in activation in higher-level perceptual regions (fusiform and parahippocampal regions specialised for face and place encoding, respectively) by target-congruent lexical cues has only been reported once (Puri et al., 2009), while several studies have reported no effects of lexical cues on perceptual processes (Eger et al., 2007; Francken, Kok, et al., 2015). However, no studies using lexical cues reported changes in BOLD signal in lower-level visual areas. The lack of lower-level perceptual engagement found in fMRI studies stands in contrast to M/EEG studies, which reported lexical ability to modify perceptual processes as early as around 100ms post stimulus presentation (Boutonnet & Lupyan, 2015; Hirschfeld et al., 2011; Landau et al., 2010; Noorman et al., 2018), suggesting that the fMRI studies might not be looking at the right neural signatures of this interaction in early visual areas. Looking at the effects of non-linguistic top-down influences on perception, Kok et al. (2012, 2014) found that top-down expectations sharpen the neural representation of expected stimuli in early visual cortices by simultaneously dampening activations coding for unexpected stimuli and amplifying activation coding for expected stimuli (Kok et al., 2012; Kok & de Lange, 2014). This kind of modulation is not usually detectable with the standard univariate approaches, given that they rely on averaging procedures to examine the overall increase in activation within a region, by looking at voxels in isolation rather than a shift in the pattern of activation as a result of the joint analysis across voxels.

Effects of Language on General Motion Perception

In the experiment with the RDM task (up/down motion discrimination), we wanted to examine the effects of lexical motion cues on general motion perception. To that end, we used thresholded random dot motion (RDM) stimuli with a distractor (PLF figure) in the middle.

Following the same logic as in the PLF task, and building upon previous studies, we expected general motion cues, directionally congruent with the RDM target, to improve accuracy on the general motion discrimination task. Instead, we found that both congruent and incongruent biological motion cues (the distractor cues in this experiment) led to higher accuracies in general motion discrimination than target-focusing congruent general motion cues. It could be that this surprising result was observed because the distractor figure in the middle of the RDM aperture provided a reference point against which the directionality of the movement around it could be judged, thus leading to better motion direction discrimination. Alternatively, we could be observing some kind of behavioural manifestation of repetition interference, reported in several behavioural studies looking at spatial and directional cueing (Estes et al., 2008; Richardson et al., 2003). However, several studies using specifically RDM stimuli with directionality judgement have observed facilitatory effects of directionally congruent words on RDM perception (Francken, Kok et al., 2015; Meteyard et al., 2007; Pavan et al., 2013). Interestingly, the attentional effect of biological motion cues towards the PLF distractor was observed even though the task drew the focus toward the RDM target, suggesting that even when the task is unrelated to the cues or the visual stimulus, the cueing effect still persists.

Conversely, reaction time results confirmed our hypothesis in showing that general motion cues directionally congruent with the RDM target led to a speed up in general motion perception. This finding is in line with previous studies on general motion perception with RDM stimuli (Francken, Kok, et al., 2015; Meteyard et al., 2007; Pavan et al., 2013). The finding that biological motion distractor cues resulted in the longest latencies, further confirms our speculation that the PLF distractors in the middle of the RDM aperture might have led to higher accuracy rates because the target in the middle served as a reference point against which the directionality was spotted more easily. Any kind of interference observed in the case of directional congruence, such as the suppression of the cued features in order to process the target features would have been apparent in reaction times as well.

Interestingly, we found behavioural cueing effects for both accuracies and reaction times only for downward moving RDMs, and not for upward RDMs. The asymmetry between upward and downward motion perception has been reported before in lexical cueing as well as visual perception and saccadic movements studies (Meteyard et al., 2007; Seya et al., 2015), but the reason for this discrepancy is still largely unknown. In our study, it could be the case that downward motion is somehow more saliently represented in our visual system than upward motion, possibly because we have historically encountered downward motion more often in everyday life due to gravity. This higher saliency could lead to stronger feature encoding from the cues and therefore stronger behavioural effect on perception.

Taken together, our behavioural results show an interesting speed-accuracy trade-off, where both fastest and least accurate responses were elicited by the cues congruent with the target, whereas the slowest but most accurate responses were elicited by the distractor cues, congruent with the PLF distractor.

The region of interest analysis found that congruent general motion cues lead to an increase in activation in both left and right MT/V5 regions and early visual cortex, compared to no language cues. Specifically, the increase in activation was found for congruent general motion cues, as compared to no language cues in the case of upward motion. Additionally, the increase was also found for congruent general motion cues when compared to incongruent general motion cues (in the case of downward motion for the left and right MT/V5 region), and congruent biological motion cues (in the case of upward motion in the early visual cortex).

This finding stands in contrast to one previous attempt to examine whether motion words can modify neural activity in MT/V5 region (Francken, Kok, et al., 2015). The reason for that could be because the previous study looked at the upward and downward motion jointly, while we looked at the perception of the two directions separately, allowing us to catch the subtle differences in how susceptible the perception of these two directions is to lexical influences.

Unlike the findings from the PLF task, lexical effects on perception in the RDM task were observed bilaterally, rather than just in the language-dominant, left hemisphere. Further research will be needed to discern the conditions that might lead to left-lateralised lexical top-down effects, as opposed to bilateral perceptual modifications as a function of lexical cues.

Final Remarks

Taken the two tasks together, we have showed that language can facilitate both biological and general motion perception. Additionally, looking at the neural signatures of such effects, we observed an increase in neural activation in both higher and lower-level perceptual areas as a function of target congruent lexical cues. While this effect was left-lateralised for biological motion perception, for general motion perception it was present in both the left and right MT/V5 regions, as well as the early visual cortex. In the following part of the thesis, we examine the presence of more subtle neural signatures, such as the sharpening account of neural activation, in response to lexical top-down influence on perception.

CHAPTER 4

Supplementary Material

Table S4.1.
Whole-brain clusters (PLF task)

Contrast	Cluster size	z-stat	p-value	x	y	z	Anatomical region
Coherent PLF							
CBM > NL	672	5.56	2.83e-10	-38	-44	-21	Left Fusiform Gyrus
		4.76		-44	-54	-17	Left Fusiform Gyrus
		4.68		-34	-90	-5	Left Middle Occipital Gyrus
		4.37		-40	-76	-9	Left Inferior Occipital Gyrus
		4.2		-44	-52	-11	Left Inferior Temporal Gyrus
		4.08		-40	-86	-9	Left Inferior Occipital Gyrus
	445	4.66	1.19e-07	-58	-58	14	Left Middle Temporal Gyrus
		4.04		-50	-40	8	Left Middle Temporal Gyrus/Superior Temporal Sulcus
		3.69		-62	-46	2	Left Middle Temporal Gyrus
		3.6		-56	-46	-3	Left Middle Temporal Gyrus
		3.39		-66	-52	2	Left Middle Temporal Gyrus
	431	5.02	1.79e-07	-52	25	24	Left Inferior Frontal Gyrus (Pars Triangularis)
		4.78		-44	25	26	Left Inferior Frontal Gyrus (Pars Triangularis)
		3.52		-50	17	26	Left Inferior Frontal Gyrus (Pars Triangularis)
		3.3		-44	9	34	Left Precentral Gyrus
		3.16		-58	33	12	Left Inferior Frontal Gyrus (Pars Triangularis)
	178	4.21	0.000974	-30	-72	44	Left Inferior Parietal Lobule
		3.85		-34	-66	52	Left Superior Parietal Lobule
		3.28		-42	-74	50	Left Angular Gyrus

NL > CBM	-	-	-	-	-	-	-
CBM > IBM	-	-	-	-	-	-	-
IBM > CBM	-	-	-	-	-	-	-
CBM > CGM	635	4.8	8.42e-10	-54	-52	38	Left Inferior Parietal Lobule
		4.34		-64	-58	34	Left Angular Gyrus
		4.33		-60	-60	32	Left Angular Gyrus
		4		-44	-66	58	Left Inferior Parietal Lobule
		3.98		-54	-56	22	Left Middle Temporal Gyrus
		3.91		-50	-58	54	Left Inferior Parietal Lobule
	190	4.18	0.000646	55	-56	36	Right Angular Gyrus
		4.07		55	-58	30	Right Angular Gyrus
		3.44		57	-54	46	Right Inferior Parietal Lobule
		3.32		55	-54	54	Right Inferior Parietal Lobule
	156	4.47	0.00249	-46	47	-9	Left Medial Orbital Gyrus
		4.45		-54	43	-9	Left Medial Orbital Gyrus
	126	4.45	0.0089	-4	41	46	Left Medial Superior Frontal Gyrus
		4.4		-2	37	44	Left Medial Superior Frontal Gyrus
		3.21		-8	37	36	Left Medial Superior Frontal Gyrus
	117	4.33	0.0133	-58	-22	-15	Left Middle Temporal Gyrus
		3.99		-64	-28	-9	Left Middle Temporal Gyrus
		3.54		-60	-36	-9	Left Middle Temporal Gyrus
	111	3.8	0.0174	43	15	58	Right Middle Frontal Gyrus
		3.68		41	19	46	Right Middle Frontal Gyrus
		3.63		41	15	52	Right Middle Frontal Gyrus
		3.45		49	19	46	Right Middle Frontal Gyrus
		3.33		47	13	48	Right Middle Frontal Gyrus
CGM > CBM	-	-	-	-	-	-	-
CBM > IGM	-	-	-	-	-	-	-
IGM > CBM	-	-	-	-	-	-	-

Scrambled PLF							
CBM > NL	381	5.23	6.56e-07	-46	19	32	Left Inferior Frontal Gyrus (Pars Triangularis)
		4.72		-46	11	36	Left Inferior Frontal Gyrus (Pars Triangularis)
		4.41		-38	23	28	Left Inferior Frontal Gyrus (Pars Triangularis)
		3.87		-52	27	38	Left Middle Frontal Gyrus
		3.86		-54	27	34	Left Middle Frontal Gyrus
		3.58		-56	19	34	Left Inferior Frontal Gyrus (Pars Opercularis)
	377	4.66	7.75e-07	63	-46	42	Right Supramarginal Gyrus
		4.3		57	-50	14	Right Middle Temporal Gyrus
		4.11		57	-40	38	Right Supramarginal Gyrus
		4.09		55	-46	28	Right Supramarginal Gyrus
		3.88		67	-42	34	Right Supramarginal Gyrus
		3.87		49	-50	18	Right Middle Temporal Gyrus
	229	4.63	0.000115	-36	-60	46	Left Inferior Parietal Lobule
		4.33		-32	-54	38	Left Inferior Parietal Lobule
	164	4.58	0.00146	-60	-54	18	Left Middle Temporal Gyrus/Superior Temporal Sulcus
		4.19		-64	-56	10	Left Middle Temporal Gyrus
		3.96		-50	-52	20	Left Middle Temporal Gyrus/Superior Temporal Sulcus
	153	4.23	0.00231	-48	-50	-25	Left Inferior Temporal Gyrus
		4.23		-42	-44	-25	Left Fusiform Gyrus
		3.73		-46	-52	-13	Left Inferior Temporal Gyrus
NL > CBM	-	-	-	-	-	-	-
CBM > IBM	-	-	-	-	-	-	-
IBM > CBM	-	-	-	-	-	-	-
CBM > CGM	484	4.81	1.19e-07	63	-46	22	Right Superior Temporal Gyrus
		4.66		65	-52	18	Right Middle Temporal Gyrus
		4.32		49	-48	16	Right Middle Temporal Gyrus

		4.29		55	-44	22	Right Superior Temporal Gyrus
		4.23		49	-46	22	Right Superior Temporal Gyrus/Sulcus
		4.2		55	-50	16	Right Middle Temporal Gyrus/Superior Temporal Sulcus
	372	4.34	3.04e-06	-62	-48	32	Left Superior Temporal Gyrus/Sulcus
		4.2		-66	-46	22	Left Superior Temporal Gyrus/Sulcus
		4.01		-50	-52	18	Left Middle Temporal Gyrus
		3.95		-56	-50	22	Left Middle Temporal Gyrus
		3.87		-54	-48	18	Left Middle Temporal Gyrus
		3.81		-58	-54	20	Left Middle Temporal Gyrus
	102	3.86	0.0357	59	-32	-1	Right Middle Temporal Gyrus
		3.85		59	-28	-5	Right Middle Temporal Gyrus
		3.67		69	-26	-3	Right Middle Temporal Gyrus
		3.54		69	-34	-1	Right Middle Temporal Gyrus
	99	3.83	0.0407	-42	-56	48	Left Inferior Parietal Lobule
		3.7		-32	-60	46	Left Superior Parietal Lobule
		3.53		-32	-64	48	Left Superior Parietal Lobule
		3.52		-30	-56	42	Left Inferior Parietal Lobule
		3.48		-34	-50	48	Left Inferior Parietal Lobule
CGM > CBM	-	-	-	-	-	-	-
CBM > IGM	-	-	-	-	-	-	-
IGM > CBM	-	-	-	-	-	-	-

Table S4.2.
Whole-brain clusters (RDM task)

Contrast	Cluster size	z-stat	p-value	x	y	z	Anatomical region
Upward RDM							
CGM > NL	1229	5.07	4.95e-12	-60	-22	52	Left Inferior Parietal Lobule
		4.65		-50	-34	64	Left Postcentral Gyrus
		4.62		-54	-28	54	Left Postcentral Gyrus
		4.56		-52	-36	60	Left Postcentral Gyrus
		4.46		-44	-38	68	Left Postcentral Gyrus
		4.4		-28	-22	78	Left Precentral Gyrus
	545	4.6	1.01e-06	8	-84	26	Right Cuneus
		4.37		15	-88	40	Right Cuneus
		4.34		11	-86	42	Right Cuneus
		4.25		21	-82	52	Right Superior Parietal Lobule
		3.94		4	-80	50	Right Precuneus
		3.92		27	-86	40	Right Superior Occipital Gyrus
	229	4.22	0.00181	-22	-88	30	Left Superior Occipital Gyrus
		4.08		-16	-92	24	Left Superior Occipital Gyrus
		3.67		-8	-88	42	Left Cuneus
		3.64		-6	-92	36	Left Cuneus
		3.5		-12	-88	44	Left Superior Occipital Gyrus
	223	4.49	0.00214	-50	-62	-21	Left Cerebellum
		4.1		-36	-46	-23	Left Fusiform Gyrus
		4.04		-40	-54	-21	Left Fusiform Gyrus
		3.64		-42	-62	-21	Left Cerebellum
		3.48		-52	-70	-17	Left Inferior Occipital Gyrus
	183	4.6	0.00679	23	-54	-23	Right Cerebellum
		4.48		21	-54	-19	Right Cerebellum
		3.57		15	-60	-13	Right Cerebellum
		3.57		-2	-60	-23	Vermis
		3.55		27	-48	-25	Right Cerebellum
		3.48		33	-50	-25	Right Cerebellum

	158	4.18	0.0146	-46	-74	-11	Left Inferior Occipital Gyrus
		3.93		-48	-82	-9	Left Inferior Occipital Gyrus
		3.8		-44	-78	-17	Left Inferior Occipital Gyrus
	135	4.11	0.0303	47	-28	62	Right Postcentral Gyrus
		3.96		55	-28	58	Right Inferior Parietal Lobule
		3.92		51	-32	62	Right Postcentral Gyrus
		3.73		39	-34	72	Right Postcentral Gyrus
		3.47		41	-22	70	Right Precentral Gyrus
		3.32		35	-32	74	Right Postcentral Gyrus
NL > CGM	-	-	-	-	-	-	-
CGM > IGM	-	-	-	-	-	-	-
IGM > CGM	156	4.32	0.000513	-40	-68	48	Left Angular Gyrus
		3.81		-32	-78	48	Left Inferior Parietal Lobule
		3.44		-38	-74	42	Left Inferior Parietal Lobule
		3.38		-40	-78	46	Left Inferior Parietal Lobule
	156	4.52	0.000513	43	-60	30	Right Angular
		3.59		49	-70	30	Right Middle Occipital Gyrus
		3.56		53	-70	30	Right Angular Gyrus
		3.45		53	-64	34	Right Angular Gyrus
	153	4.31	0.000594	-52	-74	24	Left Middle Temporal Gyrus
		4.28		-56	-74	24	Left Middle Temporal Gyrus
		3.88		-46	-66	28	Left Angular Gyrus
		3.58		-58	-64	34	Left Angular Gyrus
CGM > CBM	243	4.1	1.23e-05	-28	-26	76	Right Postcentral Gyrus
		3.97		-16	-19	80	Left Paracentral Lobule
		3.96		-50	-22	64	Left Postcentral Gyrus
		3.88		-32	-38	76	Left Postcentral Gyrus
		3.82		-36	-54	66	Left Superior Parietal Lobule
		3.6		-46	-50	64	Left Inferior Parietal Lobule
CBM > CGM	-	-	-	-	-	-	-
Downward RDM							

CGM > NL	92	4.53	0.0219	-36	3	32	Left Precentral Gyrus
		3.44		-48	11	36	Left Precentral Gyrus
		3.23		-50	15	28	Left Inferior Frontal Gyrus (Pars Triangularis)
NL > CGM	80	4.72	0.0426	-4	21	6	Left Caudate Nucleus
CGM > IGM	344	4.75	6.74e-06	-8	7	58	Left Supplementary Motor Area
		4.27		6	29	26	Right Anterior Cingulum
		4		2	11	58	Right Supplementary Motor Area
		3.95		6	19	48	Right Supplementary Motor Area
		3.71		2	19	40	Right Cingulum (middle segment)
		3.68		4	29	34	Right Cingulum (middle segment)
IGM > CGM	-	-	-	-	-	-	-
CGM > IBM	138	4.32	0.000505	47	43	-1	Right Inferior Frontal Gyrus (Pars Triangularis)
		4.22		43	45	-7	Right Inferior Frontal Gyrus (Pars Orbitalis)
		4.16		49	45	-5	Right Inferior Frontal Gyrus (Pars Orbitalis)
IBM > CGM	-	-	-	-	-	-	-

CHAPTER 5

Decoding Linguistic Top-Down Effects on Perception

Abstract

Only few studies have investigated the neural correlates of lexical top-down influences on perception, with mixed results. However, those studies only considered lexical effects on perception as reflected in the mean BOLD signal changes in perceptual regions of interest, measured with univariate approaches. In this study, we use multivariate pattern analysis to examine whether lexical influence on perception is reflected in more subtle shifts in neural activation patterns. One such subtle effect – the neural sharpening account – has proposed that top-down influences on perception can lead to simultaneous upregulation of expected and downregulation of unexpected visual inputs. The generalizability of such effects to different types of top-down processes is yet to be studied. Employing inter-subject searchlight cross-classification analysis, we found that lexical effects on motion perception can be decoded from both higher and lower-level perceptual regions. Connecting decoding results with changes in the BOLD signal further showed that, when contrasted with no language cues, the classifier is picking up on the more uniform BOLD enhancement prompted by target congruent cues. However, when contrasting two different types of lexical cues, the classifier is picking up on a more subtle interplay between simultaneous neural suppression and enhancement within a cluster of interest, akin to the proposed sharpening account. Furthermore, we suggested that the sharpening effect may even be reversed when target congruent cues are contrasted with distractor cues (i.e., cues congruent with the distracting visual input on the screen). In those cases, target-congruent cues lead to neural suppression, while distractor cues may lead to neural enhancement.

In preparation as: Slivac K., Hervais-Adelman, A., Flecken M., van den Heuvel M., Hagoort P.(in prep). Decoding Linguistic Top-Down Effects on Perception.

5.1. Introduction

Linguistic cueing studies have shown that linguistic labels have the ability to modify our perceptual processes, both behaviourally and neurally. Namely, it has been shown that reading or hearing a word shortly before a visual target can increase the speed and accuracy on a target detection or discrimination task (Boutonnet & Lupyan, 2015; Francken, Kok, et al., 2015; Meteyard et al., 2007; Slivac et al., 2021). Additionally, M/EEG studies have shown that language (e.g., linguistic labels) can modify neural processes in perceptual regions usually responsible for encoding the visual representation of the labelled category (Boutonnet & Lupyan, 2015; Hirschfeld et al., 2011; Landau et al., 2010; Noorman et al., 2018). However, only a handful of studies have examined neural signatures of linguistic influence on perception with functional magnetic resonance imaging (fMRI) and the findings reported in those studies are not fully convergent.

In a visual discrimination task, Puri et al. (2009) found an increase in activation in the fusiform face area (FFA) and the parahippocampal place area (PPA) as a function of target congruent cues (word *face* or *place,* respectively). Other fMRI studies, on the other hand, have not found any proof of lexically induced changes in neural activation of perceptual regions (Francken, Kok, et al., 2015; Pirog Revill et al., 2008). The discrepancy between findings reported by behavioural and M/EEG studies on one hand, and fMRI studies on the other hand, raise two questions: can lexically induced modifications of perceptual regions be observed with the fMRI neuroimaging method, and if yes, why has the evidence of linguistic influence on perception been evasive in fMRI studies?

One possibility is that existing fMRI studies have been looking at the wrong, or rather, incomplete signatures of lexical top-down influence on visual perception. So far, lexical cueing studies with perceptual tasks (target detection or discrimination) done in an fMRI setting have hypothesized that cue-target congruence would lead to an increase in the overall mean activation levels within a region encoding the visual representation of the cued category. This hypothesis has been measured with the standard massive univariate approach, that compares neural activity independently at each voxel, and draws conclusions based on only individually significant voxels.

The univariate approach has reliably shown that, in the absence of visual stimuli, language can spontaneously increase activation in perceptual regions while people listen to stories or sentences (Huth et al., 2016; Saygin et al., 2010; Wallentin et al., 2011). However, it is unclear how this lexically induced perceptual activation then modifies how we receive and encode visual input congruent or incongruent with what has been lexically depicted. In other words, it remains unclear what happens when we intercept the lexically induced activity in perceptual regions with visual input.

Alternatively, it could be that any lexically induced changes in the perceptual regions are overpowered by the visual input, to the extent that lexical bias is only noticeable at higher, decision-making and attention levels, but not at lower perceptual levels. However, given the number of non-fMRI studies reporting behavioural and electrophysiological effects of language on perception, we believe that the more likely scenario is that researchers so far have looked at the non-optimal signatures of this effect.

Several possible activation profiles have been proposed as viable signatures of lexical top-down influence on the perception of visual targets. One possibility could be that the top-down influence evoked by lexical cues, when congruent with the visual target, will lead to an overall increase in activation in regions that respond to that target – the so-called repetition enhancement. Another possibility is that higher-level expectations, such as those evoked by linguistic cues, can dampen the overall activity in regions encoding visual targets. This phenomenon is known as repetition or expectation suppression, and proposes that the strength of activation might be increased as a function of surprise rather than expectation (de Gardelle et al., 2013; Müller et al., 2013; Summerfield et al., 2008).

Finally, a more intricate version of the two previous accounts have been proposed, arguing that there is a subtle interplay between the two phenomena – enhancement and suppression – with the overall activation in the region responding to the target being sharpened, rather than uniformly enhanced or dampened. According to the sharpening account of neural activation, the representation of expected features gets upregulated in the visual cortex, while the representation of unexpected features gets simultaneously downregulated (de Lange et al., 2018; Kok et al., 2012).

The sharpening account can be tested with multivariate approaches to the fMRI analysis. Traditionally, univariate approaches have examined the relationship between mental representations and individual voxels. While this approach has contributed immensely to our understanding of the brain organisation, linking cognitive states to mean levels of activation across a population of individually significant voxels has its limits. Multivariate approaches, such as multivariate or multivoxel pattern analysis (MVPA) and representation similarity analysis (RSA), have allowed us to advance our understanding of neural information processing, by examining patterns of activity across voxels within a region, regardless of their individual significance levels.

A multivariate approach, therefore, can reveal distinct patterns of activation as a function of experimental conditions even when that shift is not picked up by a univariate approach. This is because the sharpening effect can be accompanied by either an overall increase, no shift or even a decrease in activation within a region of interest, depending on the ratio between neurons encoding expected stimuli in any given context, compared to those encoding unexpected stimuli.

In this study, we therefore employ cross-classification in order to examine differential patterns of activation in higher and lower-level visual regions as a function of linguistic cues. Specifically, we examine whether the top-down effect of lexical cues depicting biological and general motion can be decoded in higher and lower-level perceptual regions responding to biological and general motion perception, respectively. We then look at how the information decoded by the classifier in a searchlight manner can be related to the levels of BOLD signal recorded by the univariate analysis and extracted from the searchlight clusters. In doing that, we will test whether the classifier is picking up on a general increase in activation observed with the univariate analysis or whether cue decodability is a result of a more subtle interplay between simultaneous upregulation of target relevant information and downregulation of target irrelevant information, i.e., sharpening. Finally, building upon the finding that target congruent cues in the PLF task led to an increase in the false alarm rate (i.e., reporting seeing coherent targets on trials with scrambled targets), we examine whether this false judgement happens only at the decision-making level or whether it is also visible in the visual cortex.

Present Study

This study is comprised of two different tasks involving the perception of point-light stimuli - the PLF-task and RDM-task. Both PLFs and RDMs consist of disconnected dots comprising a human figure in motion (PLFs) or an aperture with a dominant motion direction vector (RDMs), that need to be integrated in space and time in order to recognise a coherent percept. In both tasks, we presented participants with lexical cues - nouns denoting either general (e.g., *snow*, *smoke*, etc.) or biological (e.g., *walker*, *dancer*, etc.) motion, followed by a visual target. The target depicted a biological motion stimulus – a PLF engaged in an action such as walking or dancing – embedded into a general motion stimulus – an RDM aperture with upward or downward moving dots. Both cues and targets were kept the same across two tasks. The tasks were performed sequentially and counterbalanced across participants. The PLF task drew their attention (i.e., biased them) towards biological motion perception, asking participants to indicate if the PLF in the middle of the RDM aperture represents a coherent or incoherent figure in motion. The RDM task biased participants towards general motion perception, asking them to indicate if the RDM dots around the PLF stimulus are moving predominantly in an upward or downward direction. This manipulation allowed us to look at the effect of lexical cues on both biological and general motion, using the same stimuli. Additionally, a functional localiser session void of a task was employed in order to localise regions involved in the processing of point-light motion stimuli in the brain in an uncued manner (see Methods in Chapter 4 for details).

5.2. Methods

The same 30 datasets from Chapter 4 were used in this chapter.

Correspondingly, methods and stimuli are the same as in Chapter 4, and can be found described there in detail.

5.2.1. Analysis

PLF Task

In the PLF task, the participants were asked to indicate if they see a coherent or scrambled PLF target, under the influence of 5 cue categories: biological motion congruent with the PLF target, biological motion incongruent with the PLF target, general motion directionally congruent with the RDM mask, general motion directionally incongruent with the RDM mask, and no language cues.

We were interested in the effects of our lexical cues on the perception of coherent and scrambled PLF targets respectively. Our analysis therefore contrasted congruent biological motion cues (CBM) against the control no language cue (NL), to test for the effect of biological motion labels compared to no labels. To test for the effect of congruence within the biological motion cue category, biasing perception *towards* the target, we looked at the effect of CBM cues compared to incongruent biological motion cues (IBM). Finally, to test for the effect of cues biasing perception *towards* the PLF target compared to cues biasing *away* from the target and towards the RDM mask, we contrasted the effect of CBM cues to both congruent (CGM) and incongruent (IGM) general motion cues.

Searchlight MVPA

Three ROIs were used in the analysis, two higher-level perceptual regions specific to biological motion perception (in the left and right hemisphere, respectively) and one region encompassing lower-level visual cortex. The biological motion perception ROIs used in this analysis are uniquely activated by the PLF stimuli in the functional localiser session. They encompass the lateral occipito-temporal cortex (LOTC) as well as parts of the inferior and superior parietal lobules. In addition to the two biological motion regions, the lower-level visual cortex, was also included in the analysis in order to examine whether the effects of lexical cues could be decoded from early visual regions. This region represents the most restrictive region activated by all point-light stimuli during the localiser session, encompassing the calcarine sulcus, the posterior lingual gyrus, the inferior and middle superior occipital gyri, and the cuneus, bilaterally. The early visual cortex was included as an ROI for two reasons. Firstly, because this is the region commonly activated for all point-light stimuli and therefore involved in the early stages of their perception and spatio-temporal integration. Secondly, because the sharpening effect under top-down influence was largely observed in the early visual cortex (de Lange et al., 2018; Kok et al., 2016).

The pre-processing of the fMRI data was performed in the same way as for the univariate analysis in Chapter 4 (see fMRI data pre-processing section for details). FSL FEAT was used to fit voxel-wise general linear models (GLM) to each participant's (non-smoothed) data in an event-related approach. Ten stimulus regressors of interest were included in the analysis, resulting in 10 t-maps, corresponding to the 10 experimental conditions. These conditions represent each of the 5 cueing categories (CBM, IBM, CGM, IGM, NL) nested under 2 PLF coherence conditions (coherent and scrambled) compared to baseline. Beside the stimulus regressors, the design matrix also included the following confounding regressors of no interest calculated via fMRIPrep: framewise displacement, 6 standard motion parameters, 6 anatomical CompCor components, and 2 cosines, as well as motion outliers. The approach of fitting a GLM per condition rather than per trial was employed in order to achieve stability in neural representation of our experimental conditions (Etzel et al., 2013). The resulting 10 t-maps per run per subject were then used in the cross-classification analysis.

For the cross-classification analysis, the inter-subject pattern analysis (ISPA) was performed in order to determine whether the regions uniquely activated by passively viewing PLF stimuli (i.e., regions responding to biological motion perception) were modulated by lexical cues. The ISPA approach works directly at the group level by utilizing data form all available subjects in a single analysis, training the data on a set of subjects and evaluating its generalization capability on the data from remaining subjects (Wang et al., 2020). This approach is usually employed with a cross-validated scheme, that shuffles the subjects between training and testing sets until all combinations are exhausted. The average classification accuracy across the resulting cross-validation folds significantly above chance level therefore means that a multivariate effect has been identified and that it is consistent across individuals. By doing that, the ISPA approach allows us to draw inference about the population from which the group of participants was drawn (Wang et al., 2020).

The ISPA analysis was performed on 29 participants, with one participant being excluded due to an incomplete dataset. The analysis was done with the searchlight approach. For this we used the pyMVPA package with a LinearCSVM classifier, searchlight radius of 4 voxels, and an NFoldPartitioner cross-validation scheme. This means that data has been trained on 28 participants and tested on the remaining one until all partition combinations were exhausted. The resulting 29 accuracy maps (one per cross-classification fold) where then smoothed with 4mm FWHM and entered in a non-parametric permutation one-sample t-test, performed with FSL’s randomise tool. The results were obtained after 5000 permutations and corrected for multiple comparison using the threshold-free cluster enhancement (TFCE) approach (Han et al., 2019). Only clusters above the 0.95 threshold ($p <= 0.05$) and with a minimum of 20 voxels are reported. Additionally, accuracy levels from the peaks of significant clusters were extracted and reported in order to confirm that the voxels that were most informative at the group level carried information about the conditions of interest, i.e., that their mean classification accuracy across folds is higher than chance level (Etzel et al., 2013).

In consistency with the contrasts used in the univariate analysis and our research question, 8 searchlight classifications were performed per ROI (the left and right biological motion ROIs, and the early visual cortex). For coherent and scrambled conditions respectively, the following classifications were set up: CBM vs. NL (testing the effect of congruent biological motion cues compared to no language cue), CBM vs. IBM (testing the effect of congruence within biological motion perception category), CBM vs. CGM and CBM vs. IGM (testing for the effect of cues biasing perception towards the PLF target compared to congruent and incongruent cues biasing perception away from the target and towards the RDM distractor).

BOLD Signal in Searchlight Clusters

In line with the sharpening account of top-down influences on perception, the mean BOLD signal within a region of interest may not be uniformly increased as a function of fulfilled top-down expectations or relevance for the target (biological motion cues congruent with the target, in this study). Yet, the effect of expected or target relevant features may still be decodable from that evoked by unexpected or target irrelevant features in those regions. This is because, by virtue of being non-directional, the classifier can pick up on more fine-grained patterns of activation, such as simultaneous upregulations and downregulations, usually overlooked by the univariate approaches.

In this study, we therefore wanted to examine how the significant searchlight clusters relate to the mean BOLD activity. Namely, we wanted to test whether the classifier is solely picking up on the same overall increase in activation recorded by the univariate analysis, or whether the successful decoding between two cueing conditions is a result of a more subtle interplay between neural suppression and enhancement, akin to neural sharpening.

To investigate that, we extracted the BOLD signal from the significant searchlight clusters in each of the ROIs. Specifically, we extracted parameter estimates (PEs) for the conditions of interest per participant from each significant cluster that came out of the searchlight cross-classification. We then ran mixed model analyses on those PEs, with relevant cue categories as fixed effects (sum coded: 1, -1) and the by-subject intercept as random effects.

False Alarms in the Visual Cortex

The PLF task had coherent and scrambled PLF targets. The congruent biological motion cues led to a significantly higher level of false alarms, i.e., participants reported seeing the coherent PLF when the scrambled PLF was presented (for details, see behavioural accuracy results in Chapter 4). We therefore wanted to test whether the behaviourally observed false alarm responses were also reflected in the visual cortex. In other words, we wanted to test whether the erroneous reports of seeing a coherent PLFs were accompanied by the corresponding shifts in the neural activation of the visual regions as well (akin to those when the participants were actually seeing a coherent PLF), or whether the false alarms happened purely at the decision-making level.

To test that, we examined whether false alarms and correct rejections on trials with congruent biological motion cues can be decoded from the visual regions in the brain. As a reminder, both these conditions have the exact same visual targets – scrambled PLFs, and lexical cues – congruent biological motion cues, so any difference found between the two in the visual cortex cannot be ascribed to the difference in targets or cues.

We first restricted the three ROIs used in our previous analyses further by performing a searchlight classification (with the same parameters as the searchlight analyses performed earlier to test the effects of different cues) of the hits versus correct ejections in the neutral, no language condition. This was done in order to get the unbiased clusters that respond differentially to hits and correct rejections. After obtaining significant searchlight clusters, we then tested whether within those clusters the classifier could distinguish between false alarms FAs) and correct rejections (CRs) of the scrambled PLF targets when cued by CBM cues. To achieve maximum stability of training and testing sets for the classification, we first averaged out all CBM FA trials and all CBM CR trials, respectively, per run per participant. This resulted in one CBM FA map and one CR map per run per participant. We then performed inter-subject cross-classification (with the SVM classifier) for the searchlight clusters obtained from the NL (hits vs. CRs) classification in the three ROIs. Significance of the classification accuracies were tested with non-parametric permutation one-sample t-test (with 5000 permutations).

This approach allowed us to test whether regions that show different activation patterns between conditions where people truly see coherent targets compared to when they correctly recognise scrambled targets, can also differentiate between conditions where people (erroneously) think they saw coherent targets and those when they correctly see scrambled targets.

RDM Task

In the RDM task, the participants were asked to indicate if they see an upward or downward motion direction within the RDM aperture, under the influence of the following 5 cue categories: general motion congruent with the RDM target, general motion incongruent with the RDM target, biological motion cues congruent with the PLF distractor, biological motion cues incongruent with the PLF distractor, and no language cues.

We were interested in the effects of lexical cues on the perception of upward and downward RDM targets respectively. Our analysis contrasted congruent general motion cues (CGM) against the control, no language cues (NL), to test for the effect of general motion labels compared to no labels. To test for the effect of congruence within the general motion cue category, biasing perception *towards* the RDM target, we looked at the effect of CGM cues compared to incongruent general motion cues (IGM). Finally, to test for the effect of cues biasing perception *towards* the RDM target compared to cues biasing *away* from the target and towards the PLF distractor, we contrasted the effect of CGM cues to both congruent (CBM) and incongruent (IBM) biological motion cues.

Searchlight MVPA

For the RDM task, three ROIs were again used in the analysis: bilateral general motion perception regions and the early visual cortex. The general motion perception regions for the RDM task were again the same as in the univariate analysis: the left and right MT/V5 area. The early visual cortex was the same as in the PLF task.

The first level GLM fitting was performed in the same way as for the PLF task, with 10 stimulus regressors being included in the analysis, resulting in 10 t-maps, corresponding to 10 trial conditions, representing each of the 5 cueing categories (CBM, IBM, CGM, IGM, NL) nested under 2 RDM direction conditions (upward and downward) compared to baseline.

The same ISPA approach was used, this time on all 30 participants, in the same cross-validation scheme, resulting in 30 accuracy maps (one per cross-classification fold). FSL's randomise tool with the non-parametric one-sample t-test (5000 permutations) and the TFCE correction was again used for group-level inference. Again, only clusters above the 0.95 threshold ($p <= 0.05$) and above 20 voxels or more voxels are reported.

In consistency with the contrasts reported in the univariate analysis, 8 searchlight classifications were performed per ROI (left and right MT/V5, and visual ROI). Following the same logic as for the PLF task, the following classifications were set up for upward and downward motion respectively: CGM vs. NL, CGM vs. IGM, CGM vs. CBM, and CGM vs. IBM.

BOLD Signal in Searchlight Clusters

The same investigation of how the significant searchlight clusters related to the mean BOLD activity was performed as in the PLF task. This was again done to test whether the classifier is picking up on a more uniformly distributed increase in activation, also observable by the univariate analysis, or whether the successful decoding between two cueing conditions is a result of a more subtle pattern of neural suppression and enhancement, akin to neural sharpening.

To investigate that, we extracted the BOLD signal from the searchlight clusters in each of the ROIs. Specifically, we extracted parameter estimates (PEs) for the conditions of interest per participant from each significant cluster that came out of the searchlight cross-classification. We then ran mixed model analyses on those PEs, with relevant cue categories as fixed effects (sum coded: 1, -1) and the by-subject intercept as random effects.

5.3. Results

PLF Task

Searchlight MVPA

In the left biological motion ROI (see **Table 5.1.**), when preceding the coherent PLF (see **Figure 5.1.A**), the classifier could reliably distinguish between CBM and NL cueing conditions mainly along the fusiform gyrus. When the PLF target was scrambled (see **Figure 5.1.B**), the classifier could distinguish between the activation evoked by the CBM cues compared to the NL cues in the posterior portions of the middle temporal gyrus (pMTG) and the superior temporal sulcus (pSTS). Both of these clusters were stable across searchlight radii (for results with the searchlight radius of 2 and 6 voxels, see Supplementary **Table S5.1.**).

In the right biological motion ROI (see **Table 5.1.**), on trials with the coherent PLF target (see **Figure 5.2.A** and **5.2.B**), differential patterns of activation were decoded between the CBM and CGM cueing conditions in the pMTG, as well as between the CBM and IGM cueing conditions in the superior parietal lobule (SPL). For scrambled PLF targets (see **Figure 5.2.C**), the classifier could distinguish between the CBM and NL cueing conditions in the right pMTG. While the two former clusters (CBM vs. CGM and CBM vs. IGM) were stable across different searchlight radii, the latter (CBM vs. NL) cluster was absent from the analysis with the 6 voxel searchlight radius (see Supplementary **Table S5.1.**).

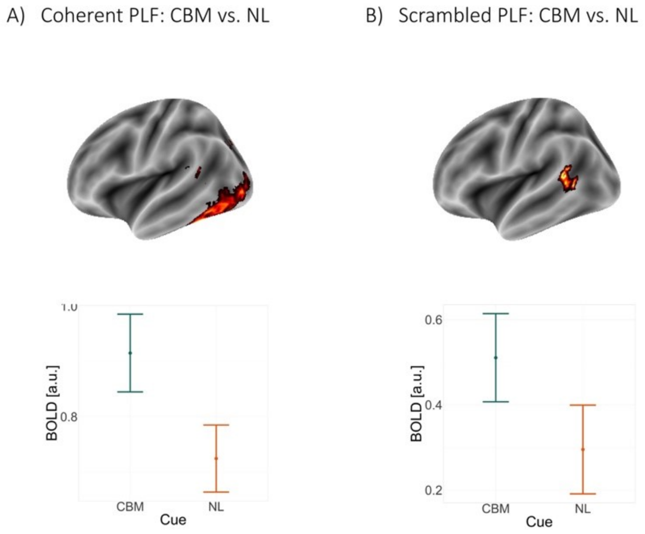
In the early visual cortex (see **Table 5.1.**), for coherent PLF trials (see **Figure 5.3.A**), the classifier could distinguish between the CBM and NL cueing conditions in the left inferior and middle occipital gyri and the lingual gyrus. For scrambled PLF trials (see **Figure 5.3.B**), the classifier could distinguish between the CBM and IBM cueing conditions in the right calcarine sulcus and the right lingual gyrus. Both clusters were consistent across searchlight radii (see Supplementary **Table S5.1.**).

Table 5.1.
Searchlight clusters (PLF task)

Target	Contrast	Clusters size	t-stats	Accuracy	x	y	z	Anatomical region
Left Biological Motion ROI								
Coherent PLF	CBM vs. NL	926	6.07	67.81%	-40	-46	-19	Left Cerebellum
			5.91	67.24%	-38	-46	-25	Left Fusiform Gyrus
			5.68	66.66%	-40	-50	-21	Left Fusiform Gyrus
			5.53	65.52%	-34	-42	-23	Left Fusiform Gyrus
			5.5	66.09%	-42	-48	-27	Left Fusiform Gyrus
			4.82	60.92%	-42	-82	-1	Left Fusiform Gyrus
Scrambled PLF	CBM vs. NL	67	4.59	61.49%	-56	-54	12	Left Middle Temporal Gyrus
			3.99	60.34%	-54	-48	8	Left Middle Temporal Sulcus
			3.98	60.34%	-52	-46	12	Left Middle Temporal Gyrus/Superior Temporal Sulcus
			3.92	60.34%	-58	-46	10	Left Middle Temporal Gyrus/Superior Temporal Sulcus
			3.58	59.20%	-50	-50	8	Left Middle Temporal Gyrus
			3.19	58.05%	-52	-52	16	Left Middle Temporal Gyrus/Superior Temporal Sulcus

Right Biological Motion ROI								
Coherent PLF	CBM vs. CGM	26	5.45	62.64%	51	-52	4	Right Middle Temporal Gyrus
			3.74	58.62%	47	-58	-1	Right Middle Temporal Gyrus
			2.98	56.89%	55	-52	2	Right Middle Temporal Gyrus
	CBM vs. IGM	48	4.45	61.49%	33	-56	66	Right Superior Parietal Lobule
			4.3	60.34%	29	-62	66	Right Superior Parietal Lobule
			3.63	58.62%	29	-62	62	Right Superior Parietal Lobule
			3.54	58.62%	29	-58	62	Right Superior Parietal Lobule
Scrambled PLF	CBM vs. NL	27	4.32	60.34%	47	-76	14	Right Middle Temporal Gyrus
Visual cortex								
Coherent PLF	CBM vs. NL	700	5.19	59.20%	-20	-84	-15	Left Lingual Gyrus
			5.11	58.05%	-32	-98	-3	Left Middle Occipital Gyrus
			4.85	58.62%	-12	-96	-7	Left Inferior Occipital Gyrus
			4.77	57.47%	-34	-96	-7	Left Middle Occipital Gyrus
			4.59	60.34%	-24	-96	-13	Left Lingual Gyrus
			4.31	59.77%	-46	-84	-7	Left Inferior Occipital Gyrus
Scrambled PLF	CBM vs. IBM	137	4.06	58.62%	11	-96	-13	Right Calcarine Sulcus
			4.04	58.62%	6	-94	-7	Right Lingual Gyrus
			4.02	59.20%	13	-86	-17	Right Calcarine Sulcus
			3.78	57.47%	11	-94	-3	Right Cerebellum
			3.5	56.90%	15	-96	-5	Right Lingual Gyrus
			3.49	57.47%	2	-94	-5	Left Calcarine Sulcus

Note: All eight contrasts were tested in all three ROIs, but only significant group-level clusters from the searchlight MVPA are reported in the table.



A) Coherent PLF: CBM vs. NL

B) Scrambled PLF: CBM vs. NL

Figure 5.1. Left biological motion ROI (LOTC and IPL) searchlights (PLF task). Group-level searchlight clusters for coherent (**A**) and scrambled (**B**) PLFs, and corresponding BOLD signal levels within those clusters. Cue names are as follows: CBM – congruent biological motion, NL – no language.

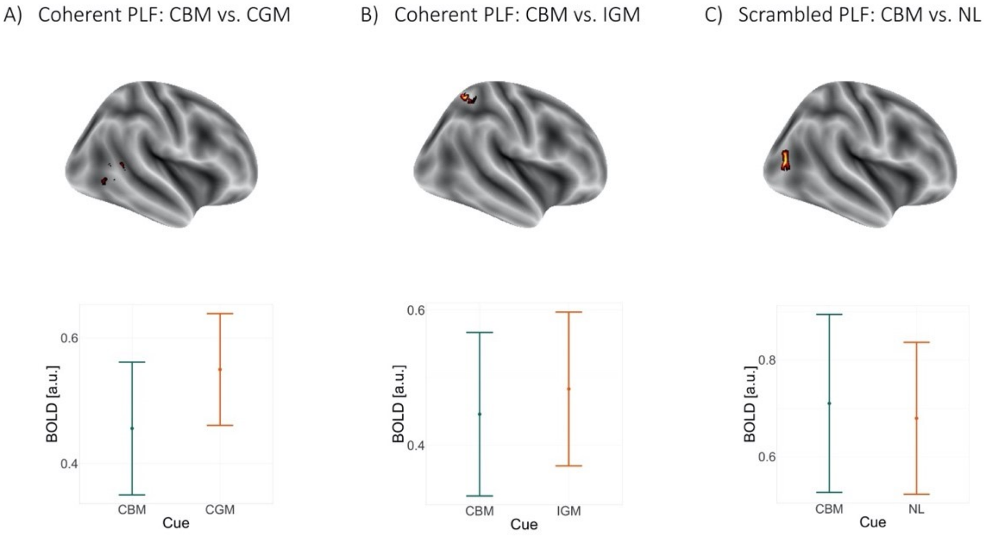


Figure 5.2. Right biological motion ROI (LOTC, IPL, SPL) searchlights (PLF task). Group-level searchlight clusters for coherent (**A**, **B**) and scrambled (**C**) PLFs, and corresponding BOLD signal within those clusters. Cue names are as follows: CBM – congruent biological motion, CGM – congruent general motion, IGM – incongruent general motion, NL – no language.

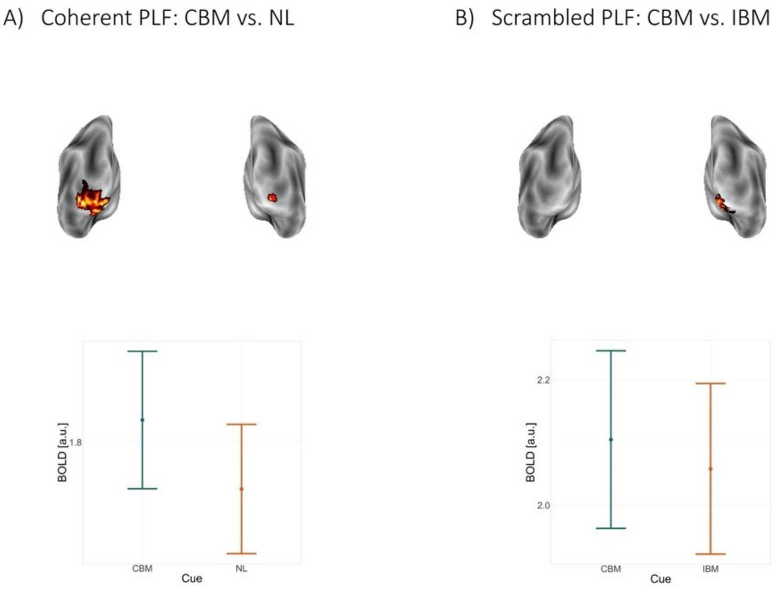


Figure 5.3. Visual cortex searchlights (PLF task). Group-level searchlight clusters for coherent (**A**) and scrambled (**B**) PLFs, and corresponding BOLD signal levels within those clusters. Cue names are as follows: CBM – congruent biological motion, IBM – incongruent biological motion, NL – no language.

BOLD Signal

In the left biological motion ROI (**Figure 5.1.**), throughout the CBM vs. NL (coherent PLFs) searchlight cluster, CBM cues evoked a more uniform increase in neural activation compared to NL cues (CBM vs. NL; estimate = 0.095, SE = 0.019, t = 5.123, p = 1.64e-05). The CBM vs. NL (scrambled PLFs) searchlight cluster showed higher activation for CBM than NL cues (CBM vs. NL; estimate = 0.108, SE =0.032, t =3.374, p = 0.002).

In the right biological motion ROI (**Figure 5.2.**), the CBM vs. CGM (coherent PLFs) searchlight cluster revealed marginally lower activation for CBM cues compared to CGM cues (CBM vs. CGM; estimate = -0.047, SE = 0.023, t = -2.019, p = 0.053). The CBM vs. IGM (coherent PLFs) searchlight cluster showed no difference in the overall activation levels (CBM vs. IGM; estimate =-0.019, SE = 0.031, t = -0.612, p = 0.545). The CBM vs. NL (scrambled PLFs) searchlight cluster showed no difference in activation for CBM compared to NL cues (CBM vs. NL, estimate = 0.015, SE = 0.026, t = 0.587, p = 0.562).

In the early visual cortex (**Figure 5.3.**), the CBM vs. NL (coherent PLFs) searchlight cluster, showed a significant increase in activation for CBM cues compared to NL cues (CBM vs. NL; estimate = 0.055, SE = 0.02, t = 2.763, p = 0.01). The CBM vs. IBM (scrambled PLFs) searchlight cluster showed no difference in activation for CBM compared to IBM cues (CBM vs. IBM; estimate = 0.023, SE = 0.028, t = 0.825, p = 0.416).

Taken together, our results show that when compared to no language cues, the classifier picked up on a more uniform increase in activation throughout the cluster, induced by congruent biological motion cues. This is also reflected in mean BOLD signal levels from the univariate analysis in both left biological motion and early visual regions. However, when compared to no language cues in the right hemisphere biological motion regions, as well as incongruent lexical cues in the right hemisphere biological motion regions and early visual regions, the classifier was picking up on a more fine-grained pattern of neural upregulation and downregulation, most likely cancelled out in the averaging procedure during the univariate analysis. When compared to general motion cues congruent with the mask in the right biological motion ROI, we even observe a dampening effect for biological motion cues congruent with the target in the corresponding searchlight cluster.

False Alarms

In the left biological motion ROI (see **Figure 5.4.A**), the NL (hits vs. correct rejections) searchlight analysis revealed significant clusters with peaks along the pMTG and the pSTS, and the inferior (IOG) and middle occipital gyri (MOG). In those clusters, the decoding accuracy of the CBM false alarms and correct rejections was above chance (mean accuracy = 58.22%, p = 0.02). In the right biological motion ROI (see **Figure 5.4.B**), the NL (hits vs. correct rejections) searchlight analysis revealed significant clusters with peaks in the pMTG and posterior inferior temporal gyrus (pITG). The decoding accuracy between the CBM false alarms and correct rejections in those clusters, however, was at chance level (mean accuracy = 53.33%, p = 0.25). In early visual cortex (see **Figure 5.4.C**), the NL (hits vs. correct rejections) searchlight analysis revealed significant clusters with peaks in the right IOG and MOG, the right cuneus, the left IOG, and the left calcarine sulcus. The decoding of the CBM false alarms and correct rejections in the searchlight clusters was above chance level (mean accuracy = 57.88%, p = 0.003).

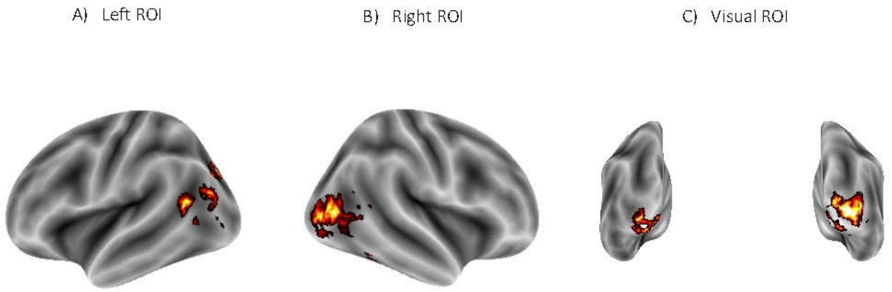


Figure 5.4. Hits vs. correct rejections. Significant searchlight clusters that carry information about hits vs. correct rejections (correctly reporting a coherent PLF vs. correctly reporting a scrambled PLF target), obtained from the no language condition. The clusters were recorded from the left (**A**) and right (**B**) biological motion regions, and the early visual cortex (**C**).

RDM Task

Searchlight MVPA

In the left and right MT/V5 regions, the searchlight analysis revealed no significant clusters for any of the cueing conditions, paired with either upward or downward motion.

In the early visual cortex (see **Table 5.2.**), the searchlight analysis revealed two significant clusters. One significant cluster was observed for the CGM vs. IBM (upward motion, **Figure 5.5.A**) cueing conditions in the right middle occipital gyrus. The searchlight results also showed a significant cluster for the CBM vs. NL (downward motion, **Figure 5.5.B**) cueing conditions in the posterior portions of the right lingual and fusiform gyri. However, neither of the clusters were consistent across searchlight radii (see Supplementary **Table S5.2.**).

Table 5.2.

Searchlight clusters (RDM task)

Target	Contrast	Clusters size	t-stats	Accuracy	x	y	z	Anatomical region
Visual cortex								
Downward RDM	CGM vs. NL	65	5.27	63.79%	25	-76	-9	Right Lingual Gyrus
			4.36	61.49%	21	-74	-11	Right Fusiform Gyrus
			4.3	60.34%	23	-76	-5	Right Cerebellum
			4.14	60.92%	27	-70	-17	Right Lingual Gyrus
			4.12	60.34%	23	-72	-15	Right Fusiform Gyrus
			3.15	57.47%	21	-82	-9	Right Lingual Gyrus
Upward RDM	CGM vs. IBM	24	5.04	62.64%	39	-86	20	Right Middle Occipital Gyrus
			4.47	60.34%	39	-82	16	Right Middle Occipital Gyrus
			4.33	59.77%	41	-86	14	Right Middle Occipital Gyrus
			3.91	59.77%	33	-84	16	Right Middle Occipital Gyrus

Note: All eight contrasts were tested in all three ROIs, but only significant group-level clusters from the searchlight MVPA are reported in the table.

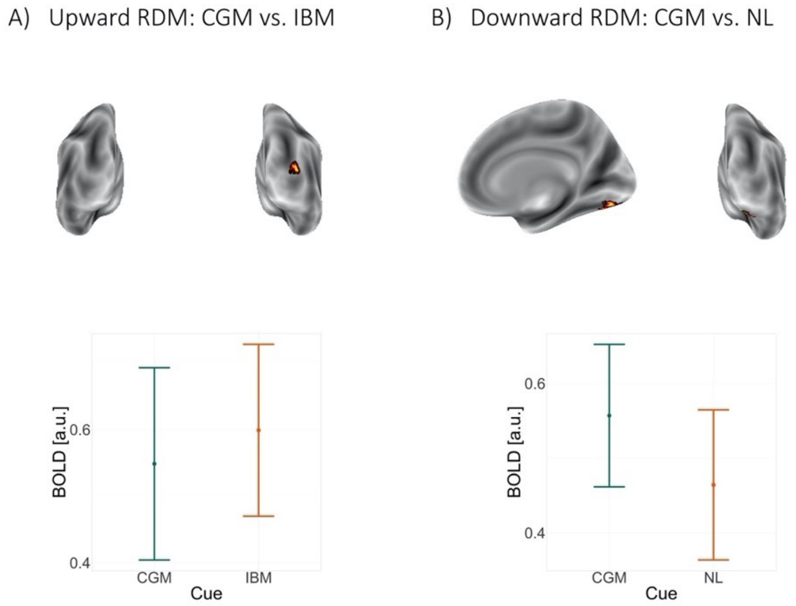


Figure 5.5. Visual cortex searchlights (RDM task). Group-level searchlight clusters for upward (**A**) and downward (**B**) motion, and corresponding BOLD signal levels in those clusters. Cue names are as follows: CGM – congruent general motion, IBM – incongruent biological motion, NL – no language.

BOLD Signal

In both left and right MT/V5 region, the overall mean activation was higher for CGM cues than NL cues (see Chapter 4 for details), however, this difference did not lead to patterns of activation distinct enough to be picked up by the classifier.

In the early visual cortex (see **Figure 5.5.**), the CGM vs. IBM (upward motion) searchlight cluster showed no difference in activation between the CGM and IBM cues (CGM vs. IBM; estimate= -0.025, SE = 0.035, t = -0.726, p = 0.473). The CGM vs. NL (downward motion) cluster showed a significantly higher activation for CGM cues than NL cues (CGM vs. NL; estimate = 0.046, SE = 0.017, t = 2.712, p = 0.011).

Taken together, the RDM results show the same patters as the PLF results. When target-congruent general motion cues were contrasted with no language cues, the classifier picked up on voxels with an increase in activation induced by the target congruent lexical cues. However, when compared to target incongruent lexical cues, the classifier was picking up on a more subtle effect of simultaneous up- and downregulation, not detectable in the mean BOLD signal levels.

5.4. Discussion

In this study, we tested whether the effects of linguistic cues can be decoded from the activation patterns in biological and general motion regions, as well as the early visual cortex. We then looked at the mean BOLD signal levels in those decoded (searchlight) clusters in order to determine whether the successful decoding was done based on the same mean enhancement picked up by the univariate analysis, or a more subtle interplay between suppression and enhancement, in line with the proposed sharpening account (de Lange et al., 2018; Kok et al., 2012; Martens & Gruber, 2012). Doing that allowed us to connect the observed neural mechanisms supporting linguistic top-down effects on perception to the more generally (non-linguistically) proposed mechanisms underlying top-down influences on perception.

Decoding Lexical Effects on Biological Motion Perception

The PLF searchlight analysis revealed significant clusters in all three perceptual regions of interest, empirically obtained from the functional localiser session.

In the left biological motion ROI, for both coherent and scrambled targets, significant searchlight clusters decoding the difference between congruent biological motion and no language cues were found in the left fusiform gyrus (coherent PLFs) and the middle temporal gyrus (scrambled PLFs). These clusters also showed a mean increase in BOLD signal for congruent biological motion cues compared to no language cues, suggesting that the classifier was picking up on an overall increase in BOLD signal throughout the cluster region. This finding is in line with the proposal put forward by the more general expectation enhancement account, proposing that expected or target relevant features upregulate the activation throughout the target encoding areas fairly uniformly (de Lange et al., 2018; Martens & Gruber, 2012).

In the right biological motion ROI, the searchlight analysis showed a significant cluster for the classification of coherent PLFs cued by congruent biological motion cues and distracting, mask-congruent general motion cues in the right posterior middle temporal gyrus. Additionally, our searchlight results revealed a significant cluster in the right superior parietal lobule for the classification of coherent PLFs cued by congruent biological motion cues and incongruent general motion cues. Finally, we find a significant searchlight cluster in the posterior right middle temporal gyrus for scrambled PLFs cued by congruent biological motion cues compared to no language cues.

However, unlike the results from the left biological motion ROI, the activation levels in the right superior parietal lobule and the middle temporal gyrus clusters revealed that the classifier was picking up on a more fine-grained interplay between neural enhancement and suppression not reflected in the mean BOLD signal levels. This activation profile may be more in line with the sharpening reports, showing a simultaneous upregulation of expected stimuli and downregulation of unexpected stimuli (de Gardelle et al., 2013; Müller et al., 2013; Summerfield et al., 2008).

Interestingly, the searchlight cluster in the right posterior middle temporal gyrus showed an overall decrease in the BOLD signal for biological motion cues congruent with the PLF target compared to distracting, general motion cues congruent with the RDM mask. Although the observed suppression effect for target relevant cues is only marginally significant, this finding warrants further inspection. Given that this pattern of activation is found only when biological motion cues are compared to general motion cues, but not incongruent general motion cues, it is not enough to hypothesise that we are observing simple dampening of target-relevant information, akin to the well documented repetition or expectation suppression effect (Barron et al., 2016; de Gardelle et al., 2013; Müller et al., 2013; Summerfield et al., 2008). Instead, it could be that the distracting, general motion cues, provided a boost in activation in the posterior middle temporal gyrus not only because they are target incongruent, but because the features encoded in them are also present on the screen – encoded in the RDM mask. In other words, it could be that our distractor cues evoked higher activation compared to biological motion cues in that cluster by virtue of matching the distractor, rather than simply mismatching the target. Simple incongruence with the target may not be enough to observe this reversal of the enhancement account, where target-incongruent cues upregulate, while target-congruent cues downregulate neural activation. Rather, the target-incongruent cues may also need to be congruent with the simultaneously presented, competing visual input.

Finally looking at the early visual cortex, our searchlight results showed a significant cluster in the left inferior and middle occipital gyri for coherent PLFs cued by congruent biological motion cues compared to no language cues. When the target PLF was scrambled, a searchlight cluster alongside the calcarine sulcus was found even for the within lexical category incongruence (for congruent vs. incongruent biological motion cues). These findings are particularly telling given the compositional nature of our visual targets – they are composed of spatially distributed dots with simple kinematics that create a more complex type of motion (i.e., biological motion) via the process of visual integration. We can therefore conclude that our biological motion stimuli already bias early visual integration processes.

Looking at the activation levels within the searchlight clusters in the visual cortex, the results show that the classifier was picking up on an overall increase in BOLD signal when congruent biological motion cues were compared with no language cues. When compared to incongruent biological motion cues, however, no difference in the mean BOLD signal levels accompanied the searchlight decoding, again suggesting a more subtle interplay between neural suppression and enhancement within that cluster, similar to that observed in the right biological motion ROI.

Particularly interesting is the finding from the congruent versus incongruent biological motion cluster, because it does not follow the neural suppression profile observed between congruent and incongruent (distracting) lexical cues in the right biological motion ROI. In addition to the hypothesis proposed above, suggesting that congruence with the mask might have played a role, along with target-incongruence, there is another explanation to keep in mind. The PLF target in this case was scrambled, while the suppression effect in the right biological motion ROI was observed for coherent PLF conditions. If this is the deciding factor in reversing the suppression effect for expected or target relevant features (i.e., those cued by congruent biological motion cues), it would not be the first time that target visibility, or rather lack thereof, was reported to reverse the suppression effect. Turk-Browne et al. (2007) showed that the repetition of low-visibility scenes led to an increase in the parahippocampal place area (PPA), whereas the opposite pattern of activation was observed for the repetition with highly visible scenes. Given that our targets were thresholded at 75% accuracy level (i.e., high-visibility level), yet compositional and noisy (i.e., masked with an RDM, and scrambled half of the time), it is possible that we are seeing this exact pattern of suppression reversal in the case of scrambled, nonsensical targets.

In addition to decoding the effects of different lexical cues on perception, we tested whether false alarms (erroneous reports of coherent figures) reflected how participants truly saw those figures. In other words, we examined whether visual areas that can distinguish between unbiased coherent and scrambled PLFs (hits vs. correct rejections in the no language cue condition), also distinguish between false alarms and correct rejections in the congruent biological motion condition. On both false alarm and correct rejection trials, the PLF target was scrambled. However, in the case of false alarms, participants reported scrambled (nonsensical) PLF targets as coherent (recognisable figures in motion). We found that in both the left biological motion ROI and the early visual cortex, the same regions that distinguish between unbiased hits and correct rejections can also distinguish between false alarms and correct rejections in the congruent biological motion cueing condition. This suggests that when participants erroneously reported seeing a coherent figure under the influence of target congruent cues, they truly did *see* the coherent figure, i.e., the action cued by the biological motion cues. In other words, false alarms may precede the decision-making level and affect the neural activity in the early visual regions as well.

Taken together, our results show an intricate pattern of activation underlying lexical top-down influences on perception. We have shown that linguistic top-down influences can modify both higher- and lower-level visual regions. Additionally, we have shown that the decoding analysis picks up on an overall expectation enhancement when target congruent cues are contrasted with cues that carry no expectations or relevance for the target (i.e., no language cues). Additionally, we find a more subtle interplay of suppression and enhancement when target congruent cues are contrasted with other lexical cues carrying competing features, and even a marginal suppression effect for coherent PLFs when target incongruent cues are also congruent with the mask. Finally, testing the effects of false alarms on the neural activity in perceptual regions, we found that false alarms may precede decision-making level and affect the neural activity in early perceptual areas.

Decoding Lexical Effects on General Motion Perception

In the RDM task, our results showed no distinct activation pattern as a function of linguistic cues for any of our classification pairings in the MT/V5 region. The univariate ROI analysis showed an overall increase in activation as function of congruent general motion cues in the left and right MT/V5 areas, revealing that this region is affected by top-down linguistic processes (see Chapter 4). However, MT/V5 region is difficult to decode, given its anatomical structure, which is why we might be seeing this absence of searchlight clusters therein.

The searchlight analysis did, however, reveal distinct patterns of activation evoked by lexical cues in the early visual cortex, although only for two classification pairs. In particular, for the perception of upward movement, we found a significant searchlight cluster for congruent general motion cues compared to incongruent biological motion distractors, in the right middle occipital gyrus. For the perception of downward motion direction, previously argued to be more salient and therefore more susceptible to top-down influences (see Chapter 4 for the discussion), we observed a significant searchlight cluster for the classification of congruent general motion cues and no language cues in the posterior right lingual and fusiform gyri. These clusters, although found with a 4 voxel searchlight radius, did not survive our cluster stability checks (the presence of the same clusters with 2 and 6 voxel radii), and should therefore be interpreted with caution.

These results show that general motion cues may already affect the activation patterns in early visual cortex. Similarly as with the PLF targets, RDM stimuli are compositional in that their principal directionality is derived by integrating spatially distributed dots into a dominant motion vector. Given this compositional nature of the RDM stimuli, we can conclude that lexical cues may already affect early stages of motion integration.

Looking at the levels of neural activation in searchlight clusters in the early visual cortex, we see the same pattern of activation as in the PLF task. Namely, comparing CGM and IBM cues showed no significant difference in the mean BOLD signal levels, suggesting that a more subtle pattern of activation was picked up by the classifier, rather than a uniform enhancement or suppression throughout the area.

However, when congruent general motion cues were compared to no language cues, we found an overall increase in the BOLD signal within the searchlight cluster. Following the same logic as in the PLF task discussion, we observe that when compared to no language cues, the classifier is picking up on a more uniform increase in the BOLD signal within the searchlight cluster, also observed with univariate methods. However, when compared to target-incongruent lexical cues, a more subtle interplay between activation enhancement and suppression, not observable with univariate approaches, is at hand.

Taken together, we have found that lexical cues can modify patterns of activation in the lower visual cortex. However, we have not found any evidence of difference in patterns of activation as a function of cues in the motion perception MT/V5 regions. Given that univariate results in Chapter 4 show an increase in the mean levels of activation in those regions for congruent general motion cues, as well as the fact that we found shifts in activation even in earlier visual areas, the lack of decoding evidence could be due to certain peculiarities in the anatomical structure of the MT/V5 region. Finally, echoing the findings from the PLF task, we found that the decoding analysis picked up on an overall enhancement in neural activation, evoked by general motion cues compared to no language (no expectation or relevance for the target) cues. However, this was not the case when the same target congruent cues were compared to competing (distracting) lexical cues (carrying competing features). This suggests that a more subtle interplay between activation enhancement and suppression, akin to the sharpening account, might be the driving force behind the decodability of different lexical cueing effects on perception.

CHAPTER 5

Supplementary Material

Cluster stability

PLF Task

In order to test the stability of the clusters reported from the 4-voxel searchlight cross-classification analysis, we performed the exact same classification analysis with searchlight radii of 2 and 6 voxels. Apart from one cluster (the CBM vs. NL cuing condition, with scrambled PLF target), all other clusters from the radius of 4 analysis were replicated, albeit with an expected variation is size.

N.B. Given that some of the results for radius of 2 voxels analysis showed a higher number of smaller clusters, it is possible that a bigger cluster from the 4 voxel analysis was broken down into several adjacent but smaller clusters, resulting in the largest cluster being smaller than in the 4-voxel radius analysis.

Table S5.1.
Searchlight clusters (radius size of 2 and 6 voxels) for the PLF task

Target	Contrast	Clusters size (radius = 2)	Cluster size (radius = 6)
Left biological motion ROI			
Coherent PLF	CBM vs. NL	509	793
Scrambled PLF	CBM vs. NL	20	148
Right biological motion ROI			
Coherent PLF	CBM vs. CGM	7	59
	CBM vs. IGM	12	52
Scrambled PLF	CBM vs. NL	10	-
Visual cortex			
Coherent PLF	CBM vs. NL	292	589
Scrambled PLF	CBM vs. IBM	6	286

Comparison of cluster sizes for searchlights of different radii – 2 vs. 6 voxels in the left and right biological motion regions and the early visual cortex.

RDM Task

Neither of the two searchlight clusters observed with the radius of 4 survived when searchlight radius was changed to either 2 or 6 voxels.

Table S5.2.
Searchlight clusters (radius size of 2 and 6 voxels) for the RDM task

Target	Contrast	Clusters size (radius = 2)	Cluster size (radius = 6)
Visual cortex			
Downward RDM	CGM vs. NL	-	-
Upward RDM	CGM vs. IBM	1	2

Comparison of cluster sizes for searchlights of different radii – 2 vs. 6 voxels in early visual cortex.

CHAPTER 6

Summary and Discussion

Linguistic influence on perception has been largely documented across studies examining hypotheses put forward by grounded cognition and cueing studies alike. However, while the general consensus by now is that neither language nor perceptual systems are isolated and mutually impenetrable modules in the brain, the exact conditions and mechanisms by which the interactions between them occur are still largely unknown.

This thesis therefore set out to examine cognitive and neural mechanisms underlying linguistic influence on perception. Outlining possible cognitive and neural processes underlying and enabling linguistic top-down effects on perception, as reported in the literature, was a necessary first step in tackling this question. In **Chapter 2**, I therefore presented an overview of the current state of knowledge regarding linguistic ability to engage perception and modify how we receive and interpret visual inputs. I connected two lines of research – grounded cognition studies and linguistic cueing studies – in order to derive cognitive and neural principles by which language can engage perception and interact with visual inputs.

Crucially, I related findings from lexical cueing studies to more general mechanisms from non-linguistic cueing studies investigating the same phenomenon – top-down influence on perception. In order to consider a full scope of neural signatures underlying linguistic effects on perception, it was necessary to look at the language system in the context of the brain. In other words, I started with the premise that even if linguistic labels exert a particularly strong top-down influence on perception compared to their non-linguistic equivalents (such as sounds or pictures), they do so by still following more general neural mechanisms and affordances, that should also be found in non-linguistic top-down processes.

In the following section, I present a summary of the findings from the lexical cueing studies presented in this thesis, and discuss those findings in the context of cognitive and neural mechanisms underlying top-down influences on perception. This is done in order to outline how lexical effects on perception fit into a wider narrative of perceptual susceptibility to top-down influences, and the implications that those findings might have for future research.

6.1. Features as the Driving Force Behind Linguistic Top-Down Influences on Perception

One of the crucial cognitive processes proposed to underly linguistic top-down influences on perception is feature extraction. Feature-based accounts of top-down influences on perception have been proposed in both non-linguistic (Galashan & Siemann, 2017; Gong & Liu, 2019; Maunsell & Treue, 2006; Saenz et al., 2002; Summerfield & Egner, 2016) and linguistic (Lupyan, 2012) literature. These accounts state that prior knowledge of the likelihood or relevance of certain features can facilitate the perception of objects with those features. Within the context of language processing, it has been proposed that linguistic labels are the optimal vehicle for the encoding of diagnostic features, by virtue of being untied to any particular instance of a category (Edmiston & Lupyan, 2015). The ability of linguistic labels to forgo unnecessary details makes the diagnostic features stand out, and not be watered down by superfluous characteristics. The feature encoding is therefore strong enough to engage perceptual regions that usually respond to visual representations of those features. In other words, language can warp or bias the perceptual system towards the labelled category, making it particularly adept at modifying how we perceive any incoming visual inputs, compared to its non-linguistic equivalents such as pictures or sounds.

While it has been theorised that features play a key role in linguistic top-down effects on perception, no studies systematically tested how the number of features encoded in cues and targets affects the cueing power of language. **Chapter 3** therefore examined the role of features, manipulated both at the level of lexical cues and visual targets, in linguistic top-down modulations of perceptual processes. I showed that the strength of lexical influence on biological motion perception increases as a function of the number of biological motion-relevant features carried by both cue and target.

The key factor enabling the testing of the feature-based account of top-down effects on perception was the use of point-light stimuli as visual targets. They allowed me to create a compositional environment within which the presence and absence of features encoded in lexical cues and visual targets could be systematically modified. For example, the fact that people rely on both form and kinematics features for the successful integration of point-light stimuli into a meaningful percept, allowed for the testing of whether and how form-only and form-and-kinematics features expressed in cues affect target perception (Experiment 1). Furthermore, I could also examine how the (in)congruence of form-and-kinematics features with the target affects target detection (Experiment 2). And finally, I could even test how removing one feature from the visual target while keeping the other constant (e.g., removing the kinematics feature while keeping the form constant) affects lexical ability to influence biological motion perception (Experiment 3).

This approach therefore allowed me to go a step further from existing studies relying on contour recognition, and systematically investigate the extent to which people rely on lexically activated categorical features for the perceptual integration of both coherent and incoherent visual stimuli. In other words, I could show that the presence and overlap of categorical features is necessary in order for the cueing effect on perception to be observed.

Furthermore, by using both coherent and scrambled point-light stimuli, I could test the effect of language on both highly visible (i.e., coherent and therefore recognisable) and ambiguous (i.e., scrambled and made nonsensical, while maintain familiar motion pattern) visual stimuli. Specifically, I showed that linguistic labels, when featurally congruent with the target, can facilitate the detection of a highly recognisable target (i.e., coherent PLF). However, if the target recognition is compromised by means of scrambling or inversion, linguistic labels induce a perceptual bias towards the labelled category, even when that bias leads to an erroneous percept.

This finding contributes to the more general debate about the nature of top-down influences on perception, i.e., whether top-down influences on perception affect discriminability (i.e., sensitivity) between stimuli or bias perception towards the label-congruent percept. Better discriminability indicates better differentiation between target and non-target as a function of target congruent cues, while bias indicates the tendency to report perceiving what has been labelled rather than the alternative (i.e., being more partial to reporting seeing the labelled category). Previously, a distinction between these two types of influences has been tied to attention and expectation processes (Summerfield & Egner, 2016) as well as target visibility (Pavan et al., 2013). According to the framework proposed by Summerfield & Egner (2016), feature-based attention (i.e., feature relevance) should optimise discriminability between two stimuli, while feature-based expectations (i.e., feature probability) should potentiate bias towards the cued category. In **Chapter 3**, however, I have observed a shift in bias both when the cues were highly predictive of the target (Experiment 1; biological motion cues were always congruent with the target) and when they were not predictive of the target (Experiment 2; there was an equal percentage of target congruent and incongruent biological motion cues). However, expectations in both of those experiments were conflated with attentional processes, so more research with orthogonalized expectation and attention processes will be needed in order to test this hypothesis further.

The account proposed by Pavan et al. (2013), on the other hand, argues that for highly visible targets, facilitation is observed in faster RTs, while for low-visibility targets the facilitation is observed in terms of an increase in discriminability. These claims are hard to verify with experiments in this thesis, because on the one hand the targets were noisy (masked) and challenging to integrate, but on the other hand, only participants performing at 75% accuracy rate or higher were tested in the experiment, indicating that they could see the targets accurately. It is therefore hard to judge whether the stimuli used in these experiments would fall into the category of low- or high-visibility, and what implications they might carry for discriminability and bias measures.

6.2. Neural Signatures of Linguistic Top-Down Influences on Perception

Neural signatures of linguistic effect on perception have only been studied with fMRI a few times, with divergent results (Francken, Kok, et al., 2015; Pirog Revill et al., 2008; Puri et al., 2009). This stands in contrast to a large number of behavioural (Lupyan & Thompson-Schill, 2012; Lupyan & Ward, 2013; Meteyard et al., 2007; Ostarek & Huettig, 2017) and M/EEG (Boutonnet & Lupyan, 2015; Hirschfeld et al., 2011; Landau et al., 2010; Noorman et al., 2018) studies showing that language can modify how people perceive visual inputs. Additionally, as mentioned in **Chapter 2**, studies on grounded semantics have shown that language can spontaneously engage visual cortex in the same feature-based manner as reported in cueing studies (Huth et al., 2016; Martin et al., 1995; Rueschemeyer et al., 2010; Saygin et al., 2010; Simmons et al., 2007; Wallentin et al., 2011). The abundant body of evidence from studies on grounded semantics makes the absence of neural evidence regarding linguistic influence on perception from cueing studies even more prominent. This discrepancy in findings begs the following question: why have lexical cueing studies done with fMRI not found evidence of linguistic influence on perception, when the evidence for such effect is abundantly reported in behavioural and M/EEG studies?

One possibility is that lexical influence reported in perceptual tasks affects only decision-making processes, while perceptual representations of visual targets remain unaffected (Firestone & Scholl, 2016; Pylyshyn, 1999). However, given that both M/EEG cueing studies and fMRI studies on grounded semantics report changes in early visual cortices as a function of language, this seems to be an unlikely scenario. A more plausible explanation may be that current fMRI studies have looked for possibly incomplete signatures of lexical influence on perception. Namely, studies so far hypothesized that target-congruent lexical cues would lead to a fairly uniform, overall increase in activation in the perceptual regions involved in the target encoding, and have used univariate fMRI analyses approaches in order to test that hypothesis. However, another possibility, suggested by a number of studies examining (non-linguistic) top-down influences on perception, is that top-down expectations and attention can evoke a shift in neural activation akin to expectation or repetition suppression (Barron et al., 2016; de Gardelle et al., 2013; Summerfield et al., 2008; Tang et al., 2018). According to the suppression accounts, expected stimuli induce weaker responses in neurons tuned for the expected input, i.e., dampen the neural activation, whereas unexpected stimuli lead to an increase in activation of neurons encoding those unexpected stimuli. Most recently, a third possibility has been proposed, arguing that top-down expectations may lead to neural sharpening, rather than an overall increase or decrease in activation. According to the sharpening account, the neurons encoding expected information get a boost in activation, while the neurons encoding unexpected information become simultaneously suppressed (de Lange et al., 2018; Kok et al., 2012; Yon et al., 2018). Depending on the ratio of neurons encoding expected and unexpected features in a region of interest as well as its size, this interplay between an activation increase and decrease may affect the mean levels of activation in that region in any which way (overall increase, overall decrease, or no significant upward or downward shift in BOLD signal). This means that while the pattern of activation within a region may have shifted as a function of lexical cues, this shift may not be noticeable with the standard univariate analysis, as it can easily be cancelled out during the averaging process. Rather, a multivariate analysis may need to be employed to examine those types of changes.

In **Chapter 4** and **5**, I therefore examined neural signatures of lexical top-down influence on perception in the light of the possible activation profiles mentioned above. In **Chapter 4**, I showed that linguistic labels can increase the mean levels of activation in higher-level perceptual regions usually engaged in biological and general motion perception, respectively. This observation is in line with the findings reported by Puri et al. (2009), who showed that linguistic labels (face vs. place) increase the overall activation in regions encoding the perceptual representation of faces (the fusiform face area) and places (the parahippocampal place area), respectively. However, these results also stand in contrast with the absence of lexical effects on motion perception regions reported by Francken, Kok, et al. (2015). In particular, an increase in the mean activation of the MT/V5 region in the RDM task was found as a function of congruent general motion cues compared to no language cues. However, guided by previous findings showing that downward motion might be more salient and therefore differently affected by cues compared to upward motion, we looked at the upward and downward motion separately. So the neural modifications as a function of up and down encoding labels may have been missed in previous studies by collapsing findings across both directions, as was done in Francken, Kok, et al. (2015). The effect of cues directionally congruent with the target was found only when compared to no language cues for upward motion, while for downward motion that effect was also observed when compared to directionally incongruent general motion cues. While more research is needed to examine the discrepancy between the perception of upward and downward motion directionality, our results suggest that the effect of lexical incongruence in the MT/V5 region may be visible only for the more salient motion vector.

Consistent with previous findings, **Chapter 4** also found the linguistic effect on perception to be left-lateralised, but only for biological motion perception: a lexically driven increase in activation was observed only in biological motion perceptual regions in the left hemisphere, but not in the right hemisphere or in the early visual cortex. The lateralisation of linguistic top-down influence on perception has been studied more directly before, with findings showing that lexical influence on perception is observed when stimuli were presented to the right visual field, but not the left (Francken, Kok et al., 2015; Zhou et al., 2010). This observation is in line with the more general finding that the Whorfian effect, the view that differences between languages bring about differences in perception and cognition of their speakers, can be observed more strongly in the left hemisphere (Drivonikou et al., 2007; Gilbert et al., 2006; Mo et al., 2011; Regier & Kay, 2009). Given the left-lateralised nature of linguistic processing, these findings suggested that the susceptibility of perceptual regions to both short-term or long-term linguistic top-down influences may be dependent on their proximity to the language network in the brain.

However, we found no such difference for general motion perception, where the effect of cues was observed both in the left and right MT/V5 region, as well as in the early visual cortex. Although we have not manipulated the presentation of stimuli with respect to left and right visual fields, this finding challenges the idea that linguistic effects on perception are dependent on cortical proximity of the perceptual region to the language network in the brain. Given the previous reports of lateralisation, further research will be needed to solve this puzzle.

Finally, in **Chapter 5**, I used multivariate pattern analysis approaches to test the presence of lexically induced sharpening and dampening in both higher- and lower-level perceptual regions. Multivariate approaches to the fMRI analysis can reveal shifts in patterns of neural activation that univariate analysis might be blind to. I found that lexical influence on perception can be decoded in both higher and lower perceptual regions, suggesting that lexical cues congruent with the target can modify the patterns of activation in perceptual regions encoding that target. Further connection between searchlight clusters and the BOLD signal levels extracted from them showed that when contrasted with no language cues, the pattern of activation picked up by the classifier usually follows the more uniform shifts in BOLD signal observed with univariate approaches. Namely, the classifier was picking up a boost in activation in target-encoding voxels as a function of target-congruent lexical cues.

However, when contrasted with other (target incongruent or distracting) linguistic cues, target congruent cues showed no significant increase or decrease in areas with highest classification accuracy. This finding is more in line with the sharpening account, suggesting that neurons encoding expected stimuli get upregulated while those encoding unexpected stimuli get suppressed. Further, I show that when the target-incongruent cues are congruent with the distractor presented on the screen, the pattern of activation can even be reversed, leading to the suppression of expected or target-relevant features and upregulation of target-incongruent features. Taken together, what we might be observing is that no language cues, which carried no featural expectations or relevance to the target, did not cue any competing features, and therefore did not evoke a strong enough surprisal or carried any competing information that could match or overpower the activity evoked by target congruent lexical cues. Distractor-congruent linguistic cues, on the other hand, did encode competing features, causing a surprisal-driven increase in activation of the neurons encoding those incongruent features, while the target congruent features were suppressed.

Additionally, these results may have been observed due to the ratio of expected and unexpected stimuli – target incongruent or distracting cues (three cueing categories) were presented in the study more often (when counting the three incongruent cueing categories together) than target-congruent cues (one cueing category). This shift in the likelihood of the target given the cue (i.e., shift in expectation) could have played a role in observing expectation enhancement when congruent cues were contrasted with no language cues, but no such enhancement when they were contrasted with the more frequent incongruent cues.

Finally, in **Chapter 5**, I have shown that erroneous percepts (i.e., false alarms) can also be decoded from perceptual regions, as compared to correct rejections. This is interesting, given that both false alarm and correct rejection trials had the exact same (scrambled) target and cues, meaning that any observed shifts in neural patterns between the two responses are due to the participants’ impression of what they saw under the influence of cues, rather than what was actually presented to them. As such, I showed that the effects of false alarms do not stay at decision-making level, but are visible in motion-specific and early visual areas as well.

In **Chapter 5**, I have therefore shown that language can modify activation in the early visual cortex as well as in higher-level, motion-specific perceptual regions, even in the absence of the overall increase in activation in those regions. To my knowledge, this is the first study that employed multivariate pattern analysis to directly examine the shifts in patterns of activation as a function of lexical top-down influences. In doing so, I illustrated the generalizability of the neural sharpening account to studies using language as cues.

Taken together, in this thesis, I have demonstrated that language can affect motion perception both behaviourally and neurally. Firstly, I have shown that the driving mechanism behind linguistic top-down influences on perception is the process of feature extraction and representation. Further, I have outlined a comprehensive profile of neural signatures underlying lexical top-down influence on perception, in both higher and lower-level visual areas. By doing so, I have highlighted the importance of looking outside of the linguistic domain, and connecting findings from neurolinguistic studies to more general cognitive and neural principles in the brain, in order to contribute to our understanding of (lexical) top-down influence on perception.

6.3. Future Directions

While I believe that studies presented in this thesis have opened up a path for further investigation of linguistic top-down influences, many questions still remain unanswered. For instance, it is still unclear under which conditions linguistic top-down influences can lead to a better discrimination between stimuli, and when they bias perception towards the cued category, even if that leads to an erroneous percept. In this thesis, I have suggested that ambiguity of the target (i.e., low visibility or noisiness) may play a role in distinguishing between these two types of influences. However, it is also likely that attention and expectation play a role in determining when each of these top-down accounts of influence occurs. Further studies are needed with orthogonalized attention and expectation processes, while controlling for target visibility, in order to answer this question.

Additionally, the interaction between sharpening and more uniform enhancement or dampening neural top-down activation profiles should be further investigated with lexical cues. It remains unclear why some cueing contrasts showed neural changes akin to the sharpening account in perceptual regions, while others led to changes resembling more steady enhancement or dampening profiles of activation. The difference may be due to a ratio of expected and unexpected stimuli, or it could have been the case that target-incongruent lexical cues induce enhancement in neural activation in line with prediction error proposals. Further research should therefore look at the effects of sharpening, enhancement and dampening, as well as their fine-grained interplay, in the context of linguistic top-down influences, while controlling for the ratio of target relevant and irrelevant cues.

Finally, the stimuli used in studies from this thesis were all compositional, dynamic, and noisy. While they allowed us to test how lexical cues affect the visual integration processes in a systematic and controlled manner, more studies need to be done with different types of visual stimuli (both high- and low-visibility), while employing multivariate pattern analyses approaches, in order to document the full spectrum of lexical effects on perception. Given that this is the first study that used multivariate pattern analysis to show that lexical cues can modify activation patterns in early and higher-level perceptual regions, future studies will reveal whether our findings stand the test of time and can be replicated.

6.4. Final Words

The findings in this thesis bring into focus how important it is to acknowledge the influence that language has on shaping the way in which we see the world. We like to believe that our view of reality is objective and unbiased, but this is far from true. As I have demonstrated in this thesis, even a brief flash of a word can change how we receive and interpret subsequent visual information. It is therefore important for us to be aware of just how easily swayed our perception can be by how we communicate, and use our words wisely and responsibly.

References

Balakrishnan, J. D. (1999). Decision processes in discrimination: Fundamental misrepresentations of signal detection theory. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(5), 1189. https://doi.org/10.1037/0096-1523.25.5.1189

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Barron, H. C., Garvert, M. M., & Behrens, T. E. J. (2016). Repetition suppression: A means to index neural representations using BOLD? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1705), 20150355. https://doi.org/10.1098/rstb.2015.0355

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

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

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

Barsalou, L. W. (2017). What does semantic tiling of the cortex tell us about semantics? *Neuropsychologia*, *105*, 18–38. https://doi.org/10.1016/j.neuropsychologia.2017.04.011

Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91. https://doi.org/10.1016/S1364-6613(02)00029-3

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious Mixed Models. *ArXiv:1506.04967 [Stat]*. http://arxiv.org/abs/1506.04967

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

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

Beintema, J. A., Georg, K., & Lappe, M. (2006). *Perception of biological motion from limited-lifetime stimuli* (Vol. 68). https://doi.org/10.3758/BF03208763

Bertenthal, B. I., & Pinto, J. (1994). Global Processing of Biological Motions. *Psychological Science*, *5*(4), 221–225. https://doi.org/10.1111/j.1467-9280.1994.tb00504.x

Bidet-Ildei, C., Sparrow, L., & Coello, Y. (2011). Reading action word affects the visual perception of biological motion. *Acta Psychologica*, *137*(3), 330–334. https://doi.org/10.1016/j.actpsy.2011.04.001

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

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765. https://doi.org/10.1037/0033-295X.113.4.700

Bonda, E., Petrides, M., Ostry, D., & Evans, A. (1996). Specific involvement of human parietal systems and the amygdala in the perception of biological motion. *Journal of Neuroscience*, *16*(11), 3737–3744.

Borghi, A. M., & Cimatti, F. (2009). Words as Tools and the Problem of Abstract Word Meanings. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *31*(31). https://escholarship.org/uc/item/58m9n8rp

Born, R. T., & Bradley, D. C. (2005). Structure and function of visual area MT. *Annual Review of Neuroscience*, *28*, 157–189. https://doi.org/10.1146/annurev.neuro.26.041002.131052

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

Boxtel, J. J. A. van, & Lu, H. (2015). Joints and their relations as critical features in action discrimination: Evidence from a classification image method. *Journal of Vision*, *15*(1), 20–20. https://doi.org/10.1167/15.1.20

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Casasanto, D. (2012). Whorfian hypothesis. *Oxford Bibliographies Online: Anthropology*. https://doi.org/10.1093/OBO/9780199766567-0058

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention During Natural Vision Warps Semantic Representation Across the Human Brain. *Nature Neuroscience*, *16*(6), 763–770. https://doi.org/10.1038/nn.3381

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

de Gardelle, V., Waszczuk, M., Egner, T., & Summerfield, C. (2013). Concurrent repetition enhancement and suppression responses in extrastriate visual cortex. *Cerebral Cortex*, *23*(9), 2235–2244. https://doi.org/10.1093/cercor/bhs211

de Koning, B. B., Wassenburg, S. I., Bos, L. T., & Van der Schoot, M. (2017). Size Does Matter: Implied Object Size is Mentally Simulated During Language Comprehension. *Discourse Processes*, *54*(7), 493–503. https://doi.org/10.1080/0163853X.2015.1119604

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764–779. https://doi.org/10.1016/j.tics.2018.06.002

Dils, A. T., & Boroditsky, L. (2010a). Visual motion aftereffect from understanding motion language. *Proceedings of the National Academy of Sciences*, *107*(37), 16396–16400. https://doi.org/10.1073/pnas.1009438107

Dils, A. T., & Boroditsky, L. (2010b). Processing unrelated language can change what you see. *Psychonomic Bulletin & Review*, *17*(6), 882–888. https://doi.org/10.3758/PBR.17.6.882

Dove, G. (2020). More than a scaffold: Language is a neuroenhancement. *Cognitive Neuropsychology*, *37*(5–6), 288–311. https://doi.org/10.1080/02643294.2019.1637338

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., & Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences*, *104*(3), 1097–1102. https://doi.org/10.1073/pnas.0610132104

Dudschig, C., Lachmair, M., de la Vega, I., Filippis, M., & Kaup, B. (2012). From top to bottom: Spatial shifts of attention caused by linguistic stimuli. *Cognitive Processing*, *13 Suppl 1*, S151-4. https://doi.org/10.1007/s10339-012-0480-x

Dudschig, C., Souman, J., Lachmair, M., Vega, I. de la, & Kaup, B. (2013). Reading "Sun" and Looking Up: The Influence of Language on Saccadic Eye Movements in the Vertical Dimension. *PLOS ONE*, *8*(2), e56872. https://doi.org/10.1371/journal.pone.0056872

Dunn, B., Kamide, Y., & Scheepers, C. (2014, July 24). *Hearing 'moon' and looking up: Word-related spatial associations facilitate saccades to congruent locations*.

Dunovan, K. E., Tremel, J. J., & Wheeler, M. E. (2014). Prior probability and feature predictability interactively bias perceptual decisions. *Neuropsychologia*, *61*, 210–221. https://doi.org/10.1016/j.neuropsychologia.2014.06.024

Edmiston, P., & Lupyan, G. (2015). What makes words special? Words as unmotivated cues. *Cognition*, *143*, 93–100. https://doi.org/10.1016/j.cognition.2015.06.008

Eger, E., Henson, R. N., Driver, J., & Dolan, R. J. (2007). Mechanisms of top-down facilitation in perception of visual objects studied by FMRI. *Cerebral Cortex (New York, N.Y.: 1991)*, *17*(9), 2123–2133. https://doi.org/10.1093/cercor/bhl119

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Estes, Z., Verges, M., & Adelman, J. S. (2015). Words, objects, and locations: Perceptual matching explains spatial interference and facilitation. *Journal of Memory and Language*, *84*, 167–189. https://doi.org/10.1016/j.jml.2015.06.002

Estes, Z., Verges, M., & Barsalou, L. W. (2008). Head Up, Foot Down: Object Words Orient Attention to the Objects' Typical Location. *Psychological Science*, *19*(2), 93–97. https://doi.org/10.1111/j.1467-9280.2008.02051.x

Etzel, J. A., Zacks, J. M., & Braver, T. S. (2013). Searchlight analysis: Promise, pitfalls, and potential. *NeuroImage*, *78*, 261–269. https://doi.org/10.1016/j.neuroimage.2013.03.041

Firestone, C., & Scholl, B. J. (2016). Cognition does not affect perception: Evaluating the evidence for 'top-down' effects. *Behavioral and Brain Sciences; New York*, *39*. http://dx.doi.org/10.1017/S0140525X15000965

Fonov, V., Evans, A., Mckinstry, R., Almli, C. R., & Collins, L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage*, *47*. https://doi.org/10.1016/S1053-8119(09)70884-5

Forder, L., & Lupyan, G. (2019). Hearing words changes color perception: Facilitation of color discrimination by verbal and visual cues. *Journal of Experimental Psychology: General*, *148*(7), 1105–1123. https://doi.org/10.1037/xge0000560

Forder, L., Taylor, O., Mankin, H., Scott, R. B., & Franklin, A. (2016). Colour Terms Affect Detection of Colour and Colour-Associated Objects Suppressed from Visual Awareness. *PLOS ONE*, *11*(3), e0152212. https://doi.org/10.1371/journal.pone.0152212

Francken, J. C., Kok, P., Hagoort, P., & de Lange, F. P. (2015). The Behavioral and Neural Effects of Language on Motion Perception. *Journal of Cognitive Neuroscience*, *27*(1), 175–184. https://doi.org/10.1162/jocn_a_00682

Francken, J. C., Meijs, E. L., Hagoort, P., van Gaal, S., & de Lange, F. P. (2015). Exploring the automaticity of language-perception interactions: Effects of attention and awareness. *Scientific Reports*, *5*. https://doi.org/10.1038/srep17725

Galashan, D., & Siemann, J. (2017). Differences and Similarities for Spatial and Feature-Based Selective Attentional Orienting. *Frontiers in Neuroscience*, *11*, 283. https://doi.org/10.3389/fnins.2017.00283

Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *249*(1325), 149–155. https://doi.org/10.1098/rspb.1992.0097

Georgeson, M. (2012). Sensory, perceptual and response biases: The criterion concept in perception. *Journal of Vision*, *12*(9), 1392–1392. https://doi.org/10.1167/12.9.1392

Gilaie-Dotan, S., Saygin, A. P., Lorenzi, L. J., Egan, R., Rees, G., & Behrmann, M. (2013). The role of human ventral visual cortex in motion perception. *Brain*, *136*(9), 2784–2798. https://doi.org/10.1093/brain/awt214

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, *103*(2), 489–494. https://doi.org/10.1073/pnas.0509868103

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience*, *14*(5), 350–363. https://doi.org/10.1038/nrn3476

Gilbert, C. D., & Sigman, M. (2007). Brain States: Top-Down Influences in Sensory Processing. *Neuron*, *54*(5), 677–696. https://doi.org/10.1016/j.neuron.2007.05.019

Gong, M., & Liu, T. (2019). Biased neural coding of feature-based attention in human brain. *BioRxiv*, 688226. https://doi.org/10.1101/688226

Gorgolewski, K. J., Burns, D. C., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). *Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python | Frontiers in Neuroinformatics*. https://www.frontiersin.org/articles/10.3389/fninf.2011.00013/full

Grèzes, J., Fonlupt, P., Bertenthal, B., Delon-Martin, C., Segebarth, C., & Decety, J. (2001). Does Perception of Biological Motion Rely on Specific Brain Regions? *NeuroImage*, *13*(5), 775–785. https://doi.org/10.1006/nimg.2000.0740

Greve, D. N. (2002) Optseq home page. Available online at: http://surfer.nmr.mgh.harvard.edu/optseq/ (Accessed on March 15th, 2019)

Grosbras, M.-H., Beaton, S., & Eickhoff, S. B. (2012). Brain regions involved in human movement perception: A quantitative voxel-based meta-analysis. *Human Brain Mapping*, *33*(2), 431–454. https://doi.org/10.1002/hbm.21222

Grossman, E. D., & Blake, R. (2002). Brain areas active during visual perception of biological motion. *Neuron*, *35*(6), 1167–1175.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of Cognitive Neuroscience*, *12*(5), 711–720.

Gumperz, J. J., & Levinson, S. C. (1991). Rethinking Linguistic Relativity. *Current Anthropology*, *32*(5), 613–623.

Han, H., Glenn, A. L., & Dawson, K. J. (2019). Evaluating Alternative Correction Methods for Multiple Comparison in Functional Neuroimaging Research. *Brain Sciences*, *9*(8), 198. https://doi.org/10.3390/brainsci9080198

Hauk, O., & Tschentscher, N. (2013). The body of evidence: What can neuroscience tell us about embodied semantics? *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00050

Hirschfeld, G., Zwitserlood, P., & Dobel, C. (2011). Effects of language comprehension on visual processing – MEG dissociates early perceptual and late N400 effects. *Brain and Language*, *116*(2), 91–96. https://doi.org/10.1016/j.bandl.2010.07.002

Hoenig, K., Sim, E.-J., Bochev, V., Herrnberger, B., & Kiefer, M. (2008). Conceptual Flexibility in the Human Brain: Dynamic Recruitment of Semantic Maps from Visual, Motor, and Motion-related Areas. *Journal of Cognitive Neuroscience*, *20*, 1799–1814. https://doi.org/10.1162/jocn.2008.20123

Hotaling, J. M., Cohen, A. L., Shiffrin, R. M., & Busemeyer, J. R. (2015). The Dilution Effect and Information Integration in Perceptual Decision Making. *PLoS ONE*, *10*(9). https://doi.org/10.1371/journal.pone.0138481

Huettig, F., Guerra, E., & Helo, A. (2020). Towards Understanding the Task Dependency of Embodied Language Processing: The Influence of Colour During Language-Vision Interactions. *Journal of Cognition*, *3*(1), 41. https://doi.org/10.5334/joc.135

Huettig, F., Olivers, C., & Hartsuiker, R. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*(2), 138–150. https://doi.org/10.1016/j.actpsy.2010.07.013

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

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

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211. https://doi.org/10.3758/BF03212378

Kerzel, D., Zarian, L., & Souto, D. (2009). Involuntary cueing effects on accuracy measures: Stimulus and task dependence. *Journal of Vision*, *9*(11), 16–16. https://doi.org/10.1167/9.11.16

Klemfuss, N., Prinzmetal, W., & Ivry, R. B. (2012). How Does Language Change Perception: A Cautionary Note. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00078

Kok, P., & de Lange, F. P. (2014). Shape Perception Simultaneously Up- and Downregulates Neural Activity in the Primary Visual Cortex. *Current Biology*, *24*(13), 1531–1535. https://doi.org/10.1016/j.cub.2014.05.042

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kok, P., van Lieshout, L. L. F., & de Lange, F. P. (2016). Local expectation violations result in global activity gain in primary visual cortex. *Scientific Reports*, *6*(1), 37706. https://doi.org/10.1038/srep37706

Kourtzi, Z., & Kanwisher, N. (2000). Activation in human MT/MST by static images with implied motion. *Journal of Cognitive Neuroscience*, *12*(1), 48–55.

Landau, A. N., Aziz-Zadeh, L., & Ivry, R. B. (2010). The Influence of Language on Perception: Listening to Sentences about Faces Affects the Perception of Faces. *The Journal of Neuroscience*, *30*(45), 15254–15261. https://doi.org/10.1523/JNEUROSCI.2046-10.2010

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *20*(7), 1434–1448. https://doi.org/10.1364/josaa.20.001434

Lenth, R. (2018). Package 'lsmeans'. *The American Statistician*, *34*(4), 216–221.

Linares, D., Aguilar-Lleyda, D., & López-Moliner, J. (2019). Decoupling sensory from decisional choice biases in perceptual decision making. *ELife*, *8*, e43994. https://doi.org/10.7554/eLife.43994

Ling, S., Liu, T., & Carrasco, M. (2009). How spatial and feature-based attention affect the gain and tuning of population responses. *Vision Research*, *49*(10), 1194–1204. https://doi.org/10.1016/j.visres.2008.05.025

Liu, T., Stevens, S. T., & Carrasco, M. (2007). Comparing the time course and efficacy of spatial and feature-based attention. *Vision Research*, *47*(1), 108–113. https://doi.org/10.1016/j.visres.2006.09.017

Louwerse, M. M. (2011). Symbol Interdependency in Symbolic and Embodied Cognition. *Topics in Cognitive Science*, *3*(2), 273–302. https://doi.org/10.1111/j.1756-8765.2010.01106.x

Lupyan, G. (2012). Linguistically Modulated Perception and Cognition: The Label-Feedback Hypothesis. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00054

Lupyan, G. (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

Lupyan, G., & Spivey, M. J. (2010a). Making the Invisible Visible: Verbal but Not Visual Cues Enhance Visual Detection. *PLoS ONE*, *5*(7), e11452. https://doi.org/10.1371/journal.pone.0011452

Lupyan, G., & Spivey, M. J. (2010b). Redundant spoken labels facilitate perception of multiple items. *Attention, Perception, & Psychophysics*, *72*(8), 2236–2253. https://doi.org/10.3758/BF03196698

Lupyan, G., & Thompson-Schill, S. L. (2012). The Evocative Power of Words: Activation of Concepts by Verbal and Nonverbal Means. *Journal of Experimental Psychology. General*, *141*(1), 170–186. https://doi.org/10.1037/a0024904

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

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

Martens, U., & Gruber, T. (2012). Sharpening and formation: Two distinct neuronal mechanisms of repetition priming. *European Journal of Neuroscience*, *36*(7), 2989–2995. https://doi.org/10.1111/j.1460-9568.2012.08222.x

Martin, A., Haxby, J. V., Lalonde, F. M., Wiggs, C. L., & Ungerleider, L. G. (1995). Discrete cortical regions associated with knowledge of color and knowledge of action. *Science; Washington*, *270*(5233), 102.

Mathôt, S., Grainger, J., & Strijkers, K. (2017). Pupillary Responses to Words That Convey a Sense of Brightness or Darkness. *Psychological Science*, *28*(8), 1116–1124. https://doi.org/10.1177/0956797617702699

Mathôt, S., Sundermann, L., & Rijn, H. van. (2019). *The effect of semantic brightness on pupil size: A replication with Dutch words* (p. 689265). https://doi.org/10.1101/689265

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

Meteyard, L., Bahrami, B., & Vigliocco, G. (2007). Motion detection and motion verbs: Language affects low-level visual perception. *Psychological Science*, *18*(11), 1007–1013. https://doi.org/10.1111/j.1467-9280.2007.02016.x

Mo, L., Xu, G., Kay, P., & Tan, L.-H. (2011). Electrophysiological evidence for the left-lateralized effect of language on preattentive categorical perception of color. *Proceedings of the National Academy of Sciences*, *108*(34), 14026–14030. https://doi.org/10.1073/pnas.1111860108

Morgan, M. J., Hole, G. J., & Glennerster, A. (1990). Biases and sensitivities in geometrical illusions. *Vision Research*, *30*(11), 1793–1810. https://doi.org/10.1016/0042-6989(90)90160-M

Müller, N. G., Strumpf, H., Scholz, M., Baier, B., & Melloni, L. (2013). Repetition suppression versus enhancement—It's quantity that matters. *Cerebral Cortex (New York, N.Y.: 1991)*, *23*(2), 315–322. https://doi.org/10.1093/cercor/bhs009

Mumford, J. (2007). A guide to calculating percent change with featquery. *Unpublished Tech Report In: http://mumford. bol. ucla. edu/perchange_guide. pdf*.

Nijhof, A. D., & Willems, R. M. (2015). Simulating Fiction: Individual Differences in Literature Comprehension Revealed with fMRI. *PLOS ONE*, *10*(2), e0116492. https://doi.org/10.1371/journal.pone.0116492

Nisbett, R. E., Zukier, H., & Lemley, R. E. (1981). The dilution effect: Nondiagnostic information weakens the implications of diagnostic information. *Cognitive Psychology*, *13*(2), 248–277. https://doi.org/10.1016/0010-0285(81)90010-4

Noorman, S., Neville, D. A., & Simanova, I. (2018). Words affect visual perception by activating object shape representations. *Scientific Reports*, *8*(1), 14156. https://doi.org/10.1038/s41598-018-32483-2

Ostarek, M., & Huettig, F. (2017). Spoken words can make the invisible visible—Testing the involvement of low-level visual representations in spoken word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 499–508. https://doi.org/10.1037/xhp0000313

Ostarek, M., & Vigliocco, G. (2017). Reading sky and seeing a cloud: On the relevance of events for perceptual simulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 579–590. https://doi.org/10.1037/xlm0000318

Pastukhov, A. (2017). First, you need a Gestalt: An interaction of bottom-up and top-down streams during the perception of the ambiguously rotating human walker. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-01376-1

Pavan, A., Skujevskis, M., & Baggio, G. (2013). Motion words selectively modulate direction discrimination sensitivity for threshold motion. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00134

Pavlova, M., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics*, *62*(5), 889–899. https://doi.org/10.3758/BF03212075

Pelekanos, V., & Moutoussis, K. (2011). The Effect of Language on Visual Contrast Sensitivity. *Perception*, *40*(12), 1402–1412. https://doi.org/10.1068/p7010

Pilly, P. K., & Seitz, A. R. (2009). What a difference a parameter makes: A psychophysical comparison of random dot motion algorithms. *Vision Research*, *49*(13), 1599–1612. https://doi.org/10.1016/j.visres.2009.03.019

Pinto, Y., Gaal, S. van, Lange, F. P. de, Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of Vision*, *15*(8), 13–13. https://doi.org/10.1167/15.8.13

Pirog Revill, K., Aslin, R. N., Tanenhaus, M. K., & Bavelier, D. (2008). Neural correlates of partial lexical activation. *Proceedings of the National Academy of Sciences*, *105*(35), 13111–13115. https://doi.org/10.1073/pnas.0807054105

Poirier, C., Collignon, O., DeVolder, A. G., Renier, L., Vanlierde, A., Tranduy, D., & Scheiber, C. (2005). Specific activation of the V5 brain area by auditory motion processing: An fMRI study. *Cognitive Brain Research*, *25*(3), 650–658. https://doi.org/10.1016/j.cogbrainres.2005.08.015

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. https://doi.org/10.1016/j.brainres.2009.08.085

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

Ramsey, R., & Hamilton, A. F. de C. (2010). Triangles have goals too: Understanding action representation in left aIPS. *Neuropsychologia*, *48*(9), 2773–2776. https://doi.org/10.1016/j.neuropsychologia.2010.04.028

Ratcliff, R., McKoon, G., & Verwoerd, M. (1989). A bias interpretation of facilitation in perceptual identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 378. https://doi.org/10.1037/0278-7393.15.3.378

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science*, *14*(4), 302–308. https://doi.org/10.1111/1467-9280.14431

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences*, *13*(10), 439–446. https://doi.org/10.1016/j.tics.2009.07.001

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*(5), 767–780. https://doi.org/10.1207/s15516709cog2705_4

Rueschemeyer, S.-A., Glenberg, A. M., Kaschak, M., Mueller, K., & Friederici, A. (2010). Top-Down and Bottom-Up Contributions to Understanding Sentences Describing Objects in Motion. *Frontiers in Psychology*, *1*. https://doi.org/10.3389/fpsyg.2010.00183

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience*, *5*(7), 631–632. https://doi.org/10.1038/nn876

Saygin, A. P. (2004). Point-Light Biological Motion Perception Activates Human Premotor Cortex. *Journal of Neuroscience*, *24*(27), 6181–6188. https://doi.org/10.1523/JNEUROSCI.0504-04.2004

Saygin, A. P. (2007a). Brain areas involved in biological motion perception: What is involved and what is necessary. *Journal of Vision*, *7*(9), 492–492. https://doi.org/10.1167/7.9.492

Saygin, A. P. (2007b). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain*, *130*(9), 2452–2461. https://doi.org/10.1093/brain/awm162

Saygin, A. P., McCullough, S., Alac, M., & Emmorey, K. (2010). Modulation of BOLD Response in Motion-sensitive Lateral Temporal Cortex by Real and Fictive Motion Sentences. *Journal of Cognitive Neuroscience*, *22*(11), 2480–2490. https://doi.org/10.1162/jocn.2009.21388

Seya, Y., Ishihara, M., & Imanaka, K. (2015). Up–down asymmetry in vertical induced motion and optokinetic nystagmus. *Attention, Perception, & Psychophysics*, *77*(1), 220–233. https://doi.org/10.3758/s13414-014-0734-z

Simmons, W., Ramjee, V., Beauchamp, M., McRae, K., Martin, A., & Barsalou, L. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*, 2802–2810. https://doi.org/10.1016/j.neuropsychologia.2007.05.002

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074. https://doi.org/10.1068/p281059

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

Stanfield, R. A., & Zwaan, R. A. (2001). The Effect of Implied Orientation Derived from Verbal Context on Picture Recognition. *Psychological Science*, *12*(2), 153–156. https://doi.org/10.1111/1467-9280.00326

Stein, T., & Peelen, M. V. (20151012). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, *144*(6), 1089. https://doi.org/10.1037/xge0000109

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Summerfield, C., & Egner, T. (2016). Feature-based attention and feature-based expectation. *Trends in Cognitive Sciences*, *20*(6), 401–404. https://doi.org/10.1016/j.tics.2016.03.008

Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, *11*(9), 1004–1006. https://doi.org/10.1038/nn.2163

Tang, M. F., Smout, C. A., Arabzadeh, E., & Mattingley, J. B. (2018). Prediction error and repetition suppression have distinct effects on neural representations of visual information. *ELife*, *7*. https://doi.org/10.7554/eLife.33123

Theeuwes, J. (2013). Feature-based attention: It is all bottom-up priming. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1628). https://doi.org/10.1098/rstb.2013.0055

Thirkettle, M., Benton, C. P., & Scott-Samuel, N. E. (2009). Contributions of form, motion and task to biological motion perception. *Journal of Vision*, *9*(3), 28–28. https://doi.org/10.1167/9.3.28

Thornton, I. M., Pinto, J., & Shiffrar, M. (1998). The Visual Perception of Human Locomotion. *Cognitive Neuropsychology*, *15*(6–8), 535–552. https://doi.org/10.1080/026432998381014

Tillas, A. (2015). Language as grist to the mill of cognition. *Cognitive Processing*, *16*(3), 219–243. https://doi.org/10.1007/s10339-015-0656-2

Tootell, R. B., Reppas, J. B., Kwong, K. K., Malach, R., Born, R. T., Brady, T. J., Rosen, B. R., & Belliveau, J. W. (1995). Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience*, *15*(4), 3215–3230.

Turk-Browne, N., Yi, D.-J., Leber, A., & Chun, M. (2007). Visual Quality Determines the Direction of Neural Repetition Effects. *Cerebral Cortex*, *17*(2), 425–433. https://doi.org/10.1093/cercor/bhj159

van Kemenade, B. M., Muggleton, N., Walsh, V., & Saygin, A. P. (2012). Effects of TMS over premotor and superior temporal cortices on biological motion perception. *Journal of Cognitive Neuroscience*, *24*(4), 896–904.

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 625–629. https://doi.org/10.3758/BF03206542

Verges, M., & Duffy, S. (2009). Spatial Representations Elicit Dual-Coding Effects in Mental Imagery. *Cognitive Science*, *33*(6), 1157–1172. https://doi.org/10.1111/j.1551-6709.2009.01038.x

Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition*, *27*, 62–75. https://doi.org/10.1016/j.concog.2014.04.007

Wallentin, M., Nielsen, A. H., Vuust, P., Dohn, A., Roepstorff, A., & Lund, T. E. (2011). BOLD response to motion verbs in left posterior middle temporal gyrus during story comprehension. *Brain and Language*, *119*(3), 221–225. https://doi.org/10.1016/j.bandl.2011.04.006

Wallentin, M., Weed, E., Østergaard, L., Mouridsen, K., & Roepstorff, A. (2008). Accessing the mental space—Spatial working memory processes for language and vision overlap in precuneus. *Human Brain Mapping*, *29*, 524–532. https://doi.org/10.1002/hbm.20413

Wang, Q., Cagna, B., Chaminade, T., & Takerkart, S. (2020). Inter-subject pattern analysis: A straightforward and powerful scheme for group-level MVPA. *NeuroImage*, *204*, 116205. https://doi.org/10.1016/j.neuroimage.2019.116205

Watson, A. B., & Pelli, D. G. (1983). Quest: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*(2), 113–120. https://doi.org/10.3758/BF03202828

Witt, J. K., Taylor, J. E. T., Sugovic, M., & Wixted, J. T. (2015). Signal Detection Measures Cannot Distinguish Perceptual Biases from Response Biases. *Perception*, *44*(3), 289–300. https://doi.org/10.1068/p7908

Yee, E., Ahmed, S. Z., & Thompson-Schill, S. L. (2012). Colorless Green Ideas (Can) Prime Furiously. *Psychological Science*, *23*(4), 364–369. https://doi.org/10.1177/0956797611430691

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027. https://doi.org/10.3758/s13423-015-0948-7

Yon, D., Gilbert, S. J., de Lange, F. P., & Press, C. (2018). Action sharpens sensory representations of expected outcomes. *Nature Communications*, *9*(1), 4288. https://doi.org/10.1038/s41467-018-06752-7

Zhang, K., & Sejnowski, T. J. (1999). Neuronal Tuning: To Sharpen or Broaden? *Neural Computation*, *11*(1), 75–84. https://doi.org/10.1162/089976699300016809

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

Zhou, K., Mo, L., Kay, P., Kwok, V. P. Y., Ip, T. N. M., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Sciences*, *107*(22), 9974–9978. https://doi.org/10.1073/pnas.1005669107

Zwaan, R. A., Stanfield, R. A., & Yaxley, R. H. (2002). Language Comprehenders Mentally Represent the Shapes of Objects. *Psychological Science*, *13*(2), 168–171. https://doi.org/10.1111/1467-9280.00430

English Summary

Language can affect what we see and how we interpret visual input. Its influence on perception is so effective that simply seeing or hearing a word before a visual stimulus can change how we perceive that stimulus. Our current understanding of this phenomenon suggests that language, by not being tied to any particular exemplar of a category, might be uniquely capable of creating strong and stable conceptual representations. These representations are potent enough to engage perceptual regions responsible for encoding visual aspects of linguistically evoked concepts. Despite numerous studies showing this effect behaviourally, neuroimaging studies looking at neural correlates of linguistic influence on perception are still largely missing. As such, the exact processes or mechanisms underlying linguistic ability to affect perception are still unknown.

This thesis therefore investigates cognitive and neural processes that allow for language-perception interaction. I used behavioural, psychophysical and neuroimaging (fMRI) methods in order to outline cognitive and neural mechanisms underlying linguistic ability to modulate perception. Specifically, the focus of studies presented in this thesis is on two types of motion perception, biological and general, as expressed in a particular type of stimuli called point-light kinematograms. These stimuli allowed me to directly measure the effect of linguistic labels on motion perception, given that the final percept they depict requires successful integration of point-light kinematics over space and time.

In Chapter 2, I provided an overview of the current state of knowledge about linguistic influence on perception, by combining behavioural and neuroimaging findings from studies on grounded cognition and (linguistic) cueing effects. In Chapter 3, I presented three experiments examining the role of feature presence and congruence for linguistic influence on biological motion perception. The results showed that feature-based conceptual activation is at the core of linguistic top-down effects on perception. In Chapters 4 and 5, I looked at the neural correlates of linguistic top-down influences on biological and general motion perception, showing that language can modify neural activity in motion-specific perceptual regions as well as in the early visual cortex. The profile of neural activation underlying this effect is complex and varies depending on the contrast. It can look like a more uniform neural enhancement in regions encoding the visual target, when cues congruent with the target are contrasted with no language. However, when contrasted with target incongruent linguistic cues, the effect is more subtle and akin to neural sharpening, and can even lead to a more

unfirm suppression when the contrasted cue is congruent with the distractor rather than the target. Further, I showed that even cue-induced false alarms (i.e., erroneously reporting seeing a stimulus that was not presented) can be successfully decoded in the visual cortex. This suggests that linguistic bias, even when leading to an erroneous percept, happens at the perceptual level rather than the higher, decision-making level.

Taken together, in this thesis I have shown that language can influence how we encode visual inputs, both behaviourally and neurally. I have empirically tested and outlined cognitive and neural mechanisms underlying such influence, highlighting their complex and context-dependent profile. In doing that, I have contributed to our understanding of the comprehensive nature of perceptual biases evoked by linguistic top-down processes.

Nederlandse Samenvatting

Taal kan invloed hebben op wat we zien en hoe we visuele input interpreteren. De invloed van taal is zo effectief dat het je waarneming van een stimulus al kan veranderen bij het zien of horen van slechts één woord voorafgaand aan het zien van de stimulus. Ons huidige begrip van dit fenomeen suggereert dat taal, door niet gebonden te zijn aan een bepaalde categorie, op een unieke wijze in staat zou kunnen zijn om conceptuele representaties te creëren. Deze representaties zijn vervolgens sterk genoeg om perceptuele regio's te activeren die verantwoordelijk zijn voor het coderen van visuele aspecten van linguïstisch opgeroepen concepten. Ondanks dat er al talrijke studies zijn die dit effect gedragsmatig kunnen aantonen, ontbreken er nog neuro-imaging-onderzoeken waarbij er gekeken wordt naar neurale correlaten van de taalkundige invloed op perceptie. Hierdoor zijn de exacte processen of mechanismen die ten grondslag liggen aan het taalvermogen om de perceptie te beïnvloeden, nog onbekend.

Dit proefschrift onderzoekt daarom de cognitieve en neurale processen die taal-perceptie interactie mogelijk maken. Ik gebruikte gedragstaken, psychofysische methoden en neuro-imaging (fMRI) om een overzicht te krijgen van de cognitieve en neurale mechanismen die ten grondslag liggen aan het taalkundig vermogen om perceptie te beïnvloeden. Specifiek ligt de focus van de studies die in dit proefschrift worden gepresenteerd op twee soorten bewegingsperceptie. Met een bepaald type stimuli, genaamd point-light kinematograms, is er gekeken naar zowel biologische als algemene bewegingsperceptie. Met deze stimuli kon ik het effect van linguïstische labels op bewegingsperceptie direct meten, aangezien het uiteindelijke beeld dat ze weergeven een succesvolle integratie van point-light kinematograms in ruimte en tijd vereist.

In Hoofdstuk 2 heb ik een overzicht gegeven van de huidige kennis over taalkundige invloed op perceptie. Hiervoor heb ik bevindingen uit zowel gedragsonderzoeken als neuro-imagingonderzoeken, waarbij er gekeken is naar cognitieve en linguïstische cue-effecten, met elkaar gecombineerd. In Hoofdstuk 3 presenteerde ik drie experimenten die de rol van de aanwezigheid van bepaalde kenmerken en de congruentie voor taalkundige invloed op biologische bewegingsperceptie onderzochten. De resultaten toonden aan dat op kenmerken gebaseerde conceptuele activering de kern vormt van taalkundige top-down effecten op perceptie. In Hoofdstukken 4 en 5 heb ik gekeken naar de neurale correlaten van linguïstische top-down invloeden op biologische en algemene bewegingsperceptie. Uit dit onderzoek blijkt dat taal de neurale activiteit in bewegingsspecifieke perceptuele gebieden en in de vroege visuele cortex kan wijzigen. Het profiel van neurale activatie dat aan dit effect ten grondslag ligt, is complex en varieert afhankelijk van het contrast. Als signalen die congruent zijn met het doelwit worden gecontrasteerd zonder taal, dan kan het eruitzien als een meer uniforme neurale versterkte activiteit in regio's die coderen voor het visuele doelwit. Wanneer het echter wordt gecontrasteerd met incongruente taalkundige signalen van het doelwit, dan is het effect subtieler en verwant aan neurale verscherping. Dit kan zelfs leiden tot een meer onvaste onderdrukking van de neurale activiteit wanneer de contrasterende cue congruent is met de afleider in plaats van met het doel. Verder toonde ik aan dat zelfs door cue geïnduceerde valse alarmen (d.w.z. foutief rapporteren van het zien van een stimulus die niet werd aangeboden) met succes kunnen worden gedecodeerd in de visuele cortex. Dit suggereert dat linguïstische vooringenomenheid, zelfs wanneer dit leidt tot een foutieve waarneming, plaatsvindt op het perceptuele niveau in plaats van op het hogere besluitvormingsniveau.

Alles bij elkaar genomen heb ik in dit proefschrift aangetoond dat taal van invloed kan zijn op hoe we visuele input coderen, zowel gedragsmatig als neuraal. Ik heb de cognitieve en neurale mechanismen die aan een dergelijke invloed ten grondslag liggen empirisch getest en een overzicht gegeven van deze mechanismen waarbij ik het complexe en context afhankelijke profiel heb benadrukt. Door dit te doen, heb ik bijgedragen aan ons begrip van de veelomvattende aard van perceptuele vooroordelen die veroorzaakt worden door linguïstische top-down processen.

Acknowledgements

When I started this PhD journey, I could not have imagined how many obstacles I would have to overcome in order to finish it. At times, I had to remind myself, as well as my colleagues, that we are scientists after all, and don't believe in curses. Still, quitting was never an option, and that was largely due to the support of my supervisors, colleagues, friends and family.

Peter, you gave me the opportunity to not only pursue my PhD at the MPI, but to do it on my own terms. You allowed me to develop my research project from scratch and supported all my scientific endeavours, however complicated or far-fetched they might have seemed. Your feedback over the years has been invaluable - you somehow always manage to make people write in their own unique way, while streamlining their thoughts into a concise and coherent narrative. I will forever be grateful for the experience and opportunities you have given me.

Monique, you allowed me to pursue my research interests even when they did not quite align with yours. Very quickly I realised I was in a lucky position where my supervisor had my best interest at heart. But you were also the person I went to whenever things got really difficult, and you were always there, armed with candy and ready to listen and empathise. I truly never felt like I was alone on this crazy journey, which means more than I can express here in words.

Alexis, we didn't always see eye-to-eye on things, but you taught me how to think and express myself clearly and precisely, and that is something I will always appreciate. Thank you for the time you invested in my PhD projects and the knowledge you shared with me over the years.

Harold, you were not directly involved in my PhD, but you were the first person who gave me a chance to explore and refine my research interests and skills, right before my PhD started. You, along with **Egbert**, **Irina** and **Sybrine** showed me what a supportive and creative research environment looked like. For that, I will forever be thankful to all of you.

Carolin, **Ina** and **Tildie**, thank you for your administrative support and all the patience in the world you have for our million questions. I am convinced this whole place would collapse without you.

Kevin, thank you for keeping us all on track, and always knowing what to do or say to make all our problems disappear. Your patience, resourcefulness, and kindness is nothing short of magical.

José, **Marcel** and **Paul**, thank you for your support with all MRI related things. I know it has been a bumpy ride, but your guidance and willingness to always find solutions were immensely helpful.

My past and present NBL lab members, **Alessio, Anne, Ashley, Atsuko, Birgit, Branka, Cas, Dan, David, Fenna, Gertje, Guillermo, Hartmut, Hanna, Hatice, Iris, Jakub, Jana, Karl Magnus, Karin, Kirsten, Linda, Mante, Margot, Marvin, Micha, Miguel, Müge, Natalia, Nienke, René, Rinus, Rowan, Sara, Teun, Valeria,** thank you for making every day at work and every lab meeting fun, educational, and something to look forward to.

Maarten, thank you for being a technical wizard and suffering through my crazy long and complicated experiments with me. I can safely say none of my projects would have been possible without you.

Jan-Mathijs, you have patiently listened as I rambled about my many analyses, taken the time to think along with me, and provided useful and thorough feedback. Thank you for being so generous with your time and knowledge.

Xin, you have sacrificed so many of your weekends and evenings to be my buddy in the MRI lab – your generosity knows no bounds.

Dick, you have been the best officemate I could possibly ask for. Our countless conversations about life were so engaging that we’d often have to remind ourselves to get back to work. You went along with all my questionable ventures throughout the years, from embracing the presence of dead plants, a beanbag and a katana in the office, to kindly tolerating my ‘artistic’ attempts to make a sculpture of Lot out of an orange. Sharing an office with you was nothing short of marvellous.

Midas, thanks for all the fun hangouts and spontaneous chats at the MPI. Your brilliant sense of humour and crazy, unexpected stories made every hangout thoroughly entertaining.

Eva, you were there when I was going through the whole thesis submission process and the waiting period that followed. It meant a lot having someone to talk to about future hopes and plans near the end of our respective contracts.

Ellie, our hangouts somehow always revolved around gin and tonics, not for the faint-hearted when made by you ;) Your warm personality combined with that cheeky, British sense of humour made for such a joyful presence in and out of work. Thank you for trusting me with your little furries, Lumpia and Shrimp.

Yingying, my workplace neighbour and travel buddy! We have gone through many adventures together, that somehow always ended up with us in the middle of a desert under the scorching sun (even in places that have no deserts!). Still, there is nobody I'd rather end up in a restricted military base in search of a statue than with you. Looking forward to our next adventure, and I promise I'll never allow anyone to steal your hat again ;)

Markus and **Nadine**, thank you for always being so warm and welcoming, for all the delicious vegan meals and wonderful conversations. **Markus**, you were not only a good friend, but also such a generous and wonderful colleague, always enthusiastic about science and happy to talk about the cool paper you'd just read. **Nadine**, our first ever hangout lasted 12 hours, which speaks volumes about our friendship. You quickly became one of my closest friends and favourite people in Nijmegen. I can't wait for our next marathon hangout!

Kristijan, thank you for always having my back, for sharing your walks in the woods and dark humour with me, and for being such a great friend all around. The way you look at life is so refreshingly unpretentious, and you always find a way to uncomplicate things.

Iris, you are not only insanely resourceful and helpful, but you also make everything seem easy and doable. Thank you for letting me babysit your equally fearless and wonderful cat **Boe**.

Laura, thanks for all the shared commiserations and enthusiasm about the MRI research. **Sophie**, thanks for being my buddy during the pandemic and the best co-paranymph on several occasions.

To my students over the years, **Irene, Jelle, Mesian, Ozge** – thank you for your interest in my projects, and I hope I didn't scare you off with my rants about science, statistics and crazy long experiments. Mentoring you was a joyous and fulfilling experience.

To my pre-PhD friends, **Andreea, Daniel, Grazia, Janna, Thijs**, thank you for making my early years in Nijmegen so uplifting and happy. I will forever have fond memories of our many shared meals, conversations and holidays.

Caitlin, you were by my side during some of the hardest moments in Nijmegen before the start of my PhD. Thank you for being there during and after my surgery, and for being such a caring and selfless friend.

Julia, **Jordy**, **Nikos** and **Suhas**, thank you for being my paranymphs and supporting me through my PhD defense. All of you have been such an important presence in my life for years now, and I am so grateful and happy to go through this final stage of my PhD journey with you by my side.

Julia, we started this crazy pursuit together and now we are finishing it together, too. I am so glad I had you to share all the joys and sorrows of the PhD life with. Thank you for being there during some of my lowest moments, and for making the thesis cover so much more beautiful than I could have ever imagined.

Jordy, you are just a ray of sunshine. I have only ever seen you be kind, responsible and open-minded, in any given situation. You are one of my closest and most beloved friends and I am so incredibly grateful to have you in my life. You and **Teresa** have always made me feel so welcome and I am such a proud aunty of your fluffy, little murder-bots, **Manou**, **Sam** and **Samou**.

Nikos, our friendship started many years ago and I am still in awe of how kind, supportive and just fantastically grounded you are. You are the best mischief buddy (especially when cats are involved), and the most creative fMRI scientist I know. Between squashes and tulips, I can’t wait to see the next crazy thing you get to scan! I cherish you and our friendship more than I can say.

Suhas (ct!), thank you for being the greatest friend a person could ask for. There is not enough space on these pages to express all the ways in which you were there for me throughout the years, so much so that I almost don’t remember a time when you were not in my life. Your enthusiasm about science, willingness to share your knowledge, and creativity are nothing short of inspiring, and I can’t wait to see all the cool things you’ll do in years to come.

Phillip, you provided a steady stream of snacks, support and surprises over the past few years, and made a valiant effort to try and keep my crazy sleep habits under control. Thank you for reminding me there was more to life than work when all I could think about was work.

Tanja, Ilakica, Veki, thank you for being the best friends and home base I could ever hope for. Your love and support over the past decade meant the world to me. You are and forever will be my family. **Tanja**, I cannot imagine going through

life without you as my best friend. So many of my favourite life adventures and memories happened with you – because of you. Your brilliant mind and fun personality make every hangout and vacation wonderful, even those with broken noses and scratched faces ;) **Ilakica**, your sharp sense of humour is a work of art, your strength is inspiring, and I will forever be grateful that you and your family welcomed me into your lives so lovingly. **Veki**, watching you over the years, I am so in awe of the person you have become. You are unbelievably smart and fearless, while somehow managing to stay the kindest person I have ever met.
I can't wait to see you and your families again!

Marko, thanks for being my go-to person when life and society stops making any sense. The way you think about the world is so fantastically unique and I always learn something new after talking to you. Looking forward to exchanging many more essays about every topic under the sun with you!

Dijana, you are one of my oldest friends and by far the most open-minded, unshockable person I have ever met. Thanks for always having my back and believing I could do anything I set my mind to. Hopefully we will find ourselves in the same country sometime in the future. I promise not to talk about decision trees ;p

To my family, **Mom** and **Branko**, thank you for your unwavering support over the years. You have been there for me from the very beginning of this journey, and I could not have done any of this without you. **Mom**, words cannot express how much I love and appreciate you. Thanks for being the best parent, the best friend, and the best role model I could wish for.

Curriculum Vitae

Ksenija Slivac was born in Virovitica, Croatia in 1986. She completed her Bachelor’s degree in English and Spanish at the University of Zagreb in Croatia. She moved to the Netherlands to pursue her Master’s degree in Cognitive Neuroscience at the Radboud University Nijmegen. Upon finishing her degree, she started her PhD at the Neurobiology of Language Department of the MPI for Psycholinguistics in Nijmegen, the Netherlands, investigating the influence of language on perception. She is currently working as a postdoctoral researcher at the MPI for Psycholinguistics.

Publications

Slivac, K., Hervais-Adelman, A., Hagoort, P., & Flecken, M. (2021). Linguistic labels cue biological motion perception and misperception. Scientific Reports, 11: 17239. doi:10.1038/s41598-021-96649-1.

Misersky, J., **Slivac, K.**, Hagoort, P., & Flecken, M. (2021). The State of the Onion: Grammatical aspect modulates object representation during event comprehension. Cognition, 214: 104744. doi:10.1016/j.cognition.2021.104744.

Slivac K., Flecken M., Hagoort P.(in prep) Grounding the Label Advantage for Perception

Slivac K., Hervais-Adelman A., Flecken M., van den Heuvel M., Hagoort P.(in prep) The Effects of Language on Biological and General Motion Perception

Slivac K., Hervais-Adelman, A., Flecken M., van den Heuvel M., Hagoort P.(in prep) Decoding Linguistic Top-Down Effects on Perception