Manuscript title: The Whorfian Brain: Neuroscientific Approaches to Linguistic Relativity

Details of all co-authors, including their full names, departments, institutions, cities, and countries of residence:

Panos Athanasopoulos, Department of Linguistics and English Language, Lancaster University, Lancaster, UK, p.athanasopoulos@lancaster.ac.uk

Aina Casaponsa, Department of Linguistics and English Language, Lancaster University, Lancaster, UK, a.casaponsa@lancaster.ac.uk

Identity and complete contact details of the corresponding author, including his/her full name, postal and e-mail addresses, and telephone and fax numbers:

Panos Athanasopoulos

Department of Linguistics and English Language

County South, Lancaster University

Bailrigg, Lancaster

United Kingdom, LA1 4YL

p.athanasopoulos@lancaster.ac.uk

+44 (0)1524 593037

five to six keywords or phrases suitable for use in an index: language, perception, categorization, linguistic relativity, Whorf

Abstract

Modern approaches to the Whorfian linguistic relativity question have reframed it from one of whether language shapes our thinking or not, to one that tries to understand the factors that contribute to the extent and nature of any observable influence of language on perception. The current paper demonstrates that such understanding is significantly enhanced by moving the evidentiary basis toward a more biologically grounded empirical arena. We review recent neuroscientific evidence using a variety of methodological techniques that reveal the functional organisation and temporal distribution of the ubiquitous relationship between language and cognitive processing in the human brain.

Introduction

The linguistic relativity hypothesis, most often linked to the writings of Benjamin Lee Whorf (Whorf, 1956) holds that language influences our perception, interpretation and understanding of the world, such that our thoughts and actions are mediated by the lexical and grammatical structures made available in our language(s). In other words, our language affects our behaviour, in predictable ways. Since its formulation, the hypothesis has been a classic topic of debate in the disciplines of Psychology, Linguistics, Anthropology and Philosophy (Lucy, 1997). The last couple of decades have witnessed an exponential growth in empirical research. Indeed, bibliometric data from Bylund and Dick (2019) reveal that the citation frequency of Whorf (1956) has skyrocketed, with 86 citations per year in the 1980s; 140 citations per year in the 1990s; 277 citations per year in the 2000s; and 420 citations per year in the 2010s.

This flood of interest has also yielded much more nuanced accounts of how language may affect behaviour and perception (see e.g. Wolff & Holmes, 2011). For example, the majority of studies show that the effects of language are dynamic and flexible, such that they may be up- or down- regulated by subtle and not so subtle experimental interventions like verbal interference paradigms, visual hemifield manipulations, perceptual discriminability of stimuli, task complexity, etc. (see Athanasopoulos, Bylund & Casasanto, 2016, for a recent collection of theoretical and empirical papers). Consequently, the traditional yes/no binary way in which researchers have tried to answer the Whorfian question (either language affects cognition, or it does not), is no longer tenable. The question has become instead a series of wh-questions: when and why do language-specific lexico-grammatical elements provide a basis for individuals to make a perceptual decision? And how does such influence play out in real time during the process of perceptual integration?

Pivotal in our epistemological enrichment of these more nuanced approaches has been the shifting of the evidentiary basis to a more biologically grounded arena, in which tangible physiological evidence for language effects on perception can be obtained. Here, we review a range of behavioural and neural measures of tracking the neural correlates of linguistically mediated behaviour. By behavioural measures, we refer to paradigms that are utilised in such a way as to provide information about the neural distribution of language-derived cognitive processing in terms of indexing attention (eye-tracking studies), laterality (visual hemifield studies), and co-activation of cell assemblies (verbal interference studies). By neural measures we refer to techniques used to reveal the functional organisation (fMRI) and temporal distribution (ERPs) of language-derived cognitive processing in the human brain.

1. Neuroscientific Approaches to Linguistic Relativity: behavioural measures

1.1. Eye-tracking

Eye-tracking (measurement of eye-movements during observation) within a Whorfian context has been used in research investigating motion event cognition. Because it can track attention towards a moving stimulus (such as those typically employed in motion research, e.g. dynamic video scenes), it is better suited to study this domain than more direct measures such as ERPs, which are best at capturing processing of single events (more on this later). Eye-tracking provides a proxy picture of information flow in the brain as it is unfolding in real time, based on the assumption that the perceptual system builds up expectations and then revises them as the scene unfolds (Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995)). Motion event studies using this technique have focused on cross-linguistic differences in grammatical aspect. Research shows that speakers of [+aspect] languages, that is, languages that obligatorily encode grammatical aspect in their verbal system (e.g. Arabic, Spanish, English) do not typically mention the endpoint of an action when describing scenes depicting an agent walking towards a potential goal (e.g. two nuns are walking) whereas speakers of [-aspect] languages, that is, languages that lack obligatory grammatical aspect (e.g. German, Swedish, Afrikaans) typically include endpoints in their descriptions (e.g. two nuns walk to a house) (Athanasopoulos & Bylund, 2013; Bylund, Athanasopoulos, & Oostendorp, 2013; von Stutterheim & Nuse, 2003). Such findings from verbal description tasks show an effect of linguistic typology on what speakers choose to mention. But what sheds light on the actual process of speech planning is an eye-tracking study by von Stutterheim, Andermann, Carroll, Flecken, and Schmiedtová (2012). The researchers studied and measured two things: 1. Number of looks to the area of interest. Here the analyses showed that speakers of [–aspect] languages made a higher number of fixations on the endpoint region, when compared to speakers of [+aspect] languages. 2. Duration of looks to the area of interest. Here the analyses revealed longer fixation duration on the endpoint region for [- aspect] speakers, when compared to [+ aspect] speakers. These findings are in line with predictions stemming from the linguistic typology. Since the endpoint will not be relevant in planning what to say, speakers of [+ aspect] languages directed less visual attention to the endpoint-object in the stimulus than speakers of [- aspect] languages. The analyses also revealed that the effect of linguistic typology on visual attention is not an all-or-nothing phenomenon. Visual attention in speakers of [+ aspect] languages also was directed to the potential endpoint to some extent (albeit to a lesser extent than in speakers of [-aspect] languages). The crucial difference between populations was that [+ aspect] speakers looked at the endpoints at a later point, compared to [- aspect] speakers (cf. von Stutterheim & Carroll, 2006). It is also important to point out that the endpoint is not actually reached in these stimulus clips. So perceptually, there is nothing to indicate that attention needs to be allocated to the endpoint of the action. But [-aspect] language speakers 'jump' forward in time, both in their linguistic descriptions, and in their dynamic visual perception. In other words, it is necessary to also take into account the time course of visual attention as an additional window on motion event processing.

This observation was made poignant in the study of Papafragou, Hulbert, and Trueswell (2008) that utilized eye tracking in Greek and English speakers. The researchers utilised cross-linguistic differences in the lexical encoding of spatial frames of reference. Speakers of so-called path languages like Greek tend to focus on the path of motion by using a relevant path verb (e.g. enter, exit etc.), while they mention the manner of motion outside of the main verb (e.g. she entered the house running). Speakers of so-called manner languages like English tend to include manner of motion in the main verb (e.g. she ran into the house). In other words, the main verb in these languages combines path and manner information. Papafragou and colleagues (2008) found no differences in attention allocation between Greek- and English-speaking observers upon first exposure to a scene. Participants looked at the path and manner elements of the action to the same extent. However, further analyses revealed that participants subsequently allocated attention to aspects of the scene not encoded in their respective languages. Greek speakers focused on manner, and English speakers on path. Intuitively this is an unexpected finding, because in this case attention allocation was the opposite of what would be expected based on the typological differences between the languages. However, the authors interpreted this "reverse Whorfian effect" as a linguistic relativity effect, because attention was guided by perceptual elements that were "new" to the observer's mental construal of the scene. The author's hypothesized that during observation, the participants implicitly verbalized the scenes (although this was not a task requirement), and those verbalizations would have been in line with the typological preferences of each language (path for Greek, manner for English). Consequently, at later stages of attention allocation, the elements that captured attention were those that had not already been implicitly verbalized by participants (path of motion for English speakers, manner of motion for Greek speakers). This explanation still leaves open the question of why participants did not direct attention to the expected areas at early stages of visual attention, but the speculation that participants verbally encoded the scenes implicitly was confirmed in a subsequent study by Trueswell and Papafragou (2010), which showed that between-group differences in allocating attention to motion events were abolished under concurrent verbal interference. We now turn to this technique.

1.2. Verbal Interference

A number of studies demonstrate that when the verbal system is simultaneously engaged in a different task, the ability to rely on verbal resources for the purposes of categorical judgments is reduced. Such findings employ a dual task methodology, whereby participants have to engage language to perform a verbal task (e.g., remembering and repeating syllables or strings of digits) parallel to the nonverbal cognitive task (e.g., categorizing motion scenes). The methodology has been extensively used to test and provide positive evidence for a well-established psycholinguistic model of human memory (Baddeley, 2003). A fundamental assumption of the model is that verbal information is encoded via a phonological loop, and that visual information may also be recoded verbally, by retrieving the corresponding label. This component allows for verbal rehearsal of stimuli in short-term memory so that they can be stored in long-term memory. Dual task paradigms employing verbal interference essentially target the verbal rehearsal system (the phonological loop), so that it can no longer be used to process verbal or verbally recoded visual stimuli (such as those used in linguistic relativity studies)

One of the first studies on linguistic relativity to use the technique captures the reasoning behind using a verbal interference condition succinctly: "This condition was designed to minimize linguistic processing of the events and to decrease memory performance by loading verbal working memory during encoding." (Gennari, Sloman, Malt, & Fitch, 2002, p. 56). In other words, the absence of a language effect on motion categorisation under verbal interference constitutes evidence of the presence of language as a strategic cognitive tool for motion categorisation under no verbal interference. The authors indeed found that in a verbal interference condition, cross-linguistic differences in similarity judgements of motion clips depicting path and manner of motion between Spanish (a path language) and English (a manner language – see previous section) speakers were abolished. Similar effects were reported by researchers studying motion event cognition using cross-linguistic differences in grammatical aspect encoding as their typological template (e.g. Athanasopoulos & Bylund, 2013). More recent findings tested the effects of verbal interference in bilingual individuals. Athanasopoulos and colleagues (2015) selectively disrupted access to one or the other language by asking German-English bilingual participants to repeat strings of numbers either in English (a [+aspect] language) or in German (a [-aspect] language – see previous section). In a within-subjects design, participants shifted their similarity judgments toward the patterns associated with the undisrupted language (so they resembled the German pattern when interference was performed in English, and vice versa with German interference). Such findings not only show a strong association between a specific language and motion event cognition, but also reveal the highly malleable nature of linguistically mediated cognition: Humans are used to mediating their cognitive judgments through language almost by default, and when access to one language is disrupted they will resort to another language if they have it at their disposal, as in the case of bilingualism.

Taken together, findings from verbal interference paradigms show that language affects cognition in real time, that is, participants draw on linguistic resources as soon as they are called to make a decision on non-verbal stimuli. This phenomenon has become the basic tenet of the label-feedback hypothesis (Lupyan, 2012). Specifically, the hypothesis holds that the verbal labels/representations of non-verbal stimuli (e.g., colours, objects, motion events) are automatically activated even in contexts that do not explicitly require the use of language (e.g. non-verbal similarity judgments). A concurrent verbal interference task disrupts processing in the phonological loop, essentially dampening the online feedback between stimuli and their corresponding linguistic representations.

The assumption that this phenomenon is the evidentiary hallmark of the online role of language as a facilitator in a concurrent nonverbal task has been bolstered by studies employing training paradigms. In one such study, Athanasopoulos and Albright (2016) asked native speakers of English to match a target showing a goal-oriented motion event (e.g. a woman walking towards a house) with a [+endpoint alternate] (a woman reaching and getting inside the house) and a [-endpoint alternate] (a woman walking towards no specific endpoint). Previous studies had shown that speakers of [+aspect] languages like English tend to match the target with the [-endpoint] alternate to a greater extent than speakers of [-aspect languages] like German (see also previous section). In this study, the researchers provided feedback to the participants: In an 'English-like' condition, the [-endpoint] alternate was the correct answer, and in a 'German-like' condition, the [+endpoint] alternate was the correct answer. As expected, when no verbal interference condition was present, participants were more successful in learning the 'English-like' pattern, since that is the pattern promoted by the aspectual system in their native language. However, under verbal interference, learning was affected only in the 'English-like' condition, that is, the condition that promoted the perceptual dimension associated with the observer's native language. In the 'German-like' condition, learning steadily improved with increasing exposure to the stimuli and the feedback, to a point that was statistically similar to the English-typical pattern. This finding is not unexpected, if one takes into account the demands that dual task paradigms place on cognitive resources. It provides evidence that individuals may recruit verbal processes online for the purposes of classification more readily when the stimuli to be classified are also habitually encoded in the native language. Verbal interference disrupts this process. In the condition that is not congruent with the native language, verbal interference in English has little effect because the grammatical pattern in English for describing goal-oriented motion is not relevant to successful categorisation.

Selective effects of verbal interference of this kind are in line with studies of individuals with aphasia. For instance, a case study of a patient with anomia shows that classification judgments of coloured objects based on taxonomic/ thematic relationships not readily codable by linguistic labels remained intact (J. Davidoff & Roberson, 2004), but classification judgments of specific perceptual dimensions of stimuli such as colour, where reliance on language is paramount for the purposes of categorical judgments were impaired (Roberson, Davidoff, & Braisby, 1999). In a more systematic investigation of 12 patients with anomia, Lupyan and Mirman (2013) instructed participants to select all objects in an array that matched a specific criterion. In a ‘high-dimensional’ condition, grouping was possible on the basis of many different features (e.g., participants were told to “select all the farm animals”). In a ‘low-dimensional’ condition, grouping required attention to one specific feature while abstracting across other task-irrelevant dimensions (e.g., participants were told to “select all the green objects”). Patients with anomia performed better in the former condition, because, according to the authors, performance on tasks utilizing low-dimensional stimuli rely on more online support from language, a resource which is impaired in patients with anomia. These findings, taken together with the findings from dual task paradigms in normal populations suggest that when participants are used to relying on prior linguistic knowledge to make classification decisions, such as encoding features of stimuli like colours and, in the case of motion, the aspectual properties of motion events, the verbal mediation that would occur by default is disrupted by verbal interference. Conversely, when perceptual decisions involve stimuli that are not habitually coded through the verbal route, participants can and do utilize non-linguistic means of classification. A number of studies have taken the online warping of categorical perception by language a step further, by looking at whether the target stimulus is presented on the left or right visual field of the observer. We turn to those studies in the next section.

1.3. Lateralisation studies

Most studies employing the so-called ‘lateralized Whorf’ paradigm, where target stimuli are shown to the left or right visual hemifield, are in the domain of colour. This is perhaps not surprising as colour has been a classic test-case of the linguistic relativity hypothesis. Unlike motion, colour is a static physical construct. Everyone sees colour biologically the same way, yet different languages carve the colour spectrum in remarkably different ways, marking contrasting categorical divisions. For example, many of the world’s languages have five basic colour terms, including a term that denotes both blue and green, a so-called ‘grue’ term found in e.g. Himba (Namibia), Berinmo (Papua New Guinea), and historically in Welsh, Japanese, and Chinese. Russian, Greek, and Turkish have two separate terms for blue, one referring exclusively to darker shades, and one referring to lighter shades.

What is the importance of having a single term for ‘light blue’ or ‘dark blue’, or ‘blue-green’ for that matter? The answer is: codability, classically defined in the memory literature as "a generic term referring to the class of stimulus properties-e.g., physical, associative, linguistic-which control the ease with which stimulus items may be placed in memory" (Ellis & Shumate, 1973, p.71). In the context of linguistic relativity research, it refers to the ease and degree of agreement with which people can name a referent. The fewer words it takes to refer to something, the more codable it is. And the more codable a referent is, the more readily it is recognized, remembered, and discriminated in non-linguistic tasks. In the context of colour terminology, as early as the 1950's Brown & Lenneberg (1954) showed differences between Zuni and English speakers in colour recognition memory as a function of how codable the colours were in their respective language.

Indeed, more recent studies established that having a label for a specific colour category helps us recognise it faster and more accurately among distractors. For example, Himba speakers who have a grue term recognize and remember greenish-blue or bluish-green focal colour stimuli better than English speakers, who have better recognition memory for blue or green focal colour stimuli (Roberson, Davidoff, Davies, & Shapiro, 2005). Linguistic influence over memory-based discrimination is perhaps not surprising, given that language can mediate memory processes, as we saw in the previous sections. The seminal study of Winawer and colleagues (2007) extended Roberson's investigation of memory to online discrimination. The researchers presented Russian and English speakers with different shades of blue (arranged in triads) and asked them to spot the odd one out as quickly as they could. Some triads included a stimulus that belonged to a different lexical blue category in Russian but not in English (cross-category triads) and in some triads all blue stimuli belonged to the same lexical category in Russian and in English (within-category triads). Results showed faster cross-category responses than within-category responses, exclusively in Russian speakers. Crucially, verbal interference abolished the Russian speakers' discrimination advantage, establishing the role of language not only in tasks that involve memory, but also in online perceptual tasks.

More recent studies have taken the online language-induced warping of colour categorical perception a step further, by looking at whether the target stimulus is presented on the left or right visual field of the observer. The relevant thing to note is that language processing, in most individuals, disproportionately involves the left hemisphere of the brain. This raises an interesting possibility. Perhaps visual information presented in our right visual field and processed first by our language-dominant left hemisphere is affected by language to a greater degree than information presented in the left visual field, first perceived by the right hemisphere. A pioneering study by Gilbert, Regier, Kay, and Ivry (2006) put this hypothesis to the test. Participants were instructed to spot the odd one out from a display of 12 colour chips arranged around a central fixation cross. Participants were told to press a key on the right side of the keyboard to indicate if the different colour was to the right of the fixation cross, and a key on the left side of the keyboard to indicate if the different colour was to the left of the fixation cross. There were two different conditions in the experiment: In the within category condition, the different stimulus came from the same lexical category as the other stimuli. In the between category condition, the deviant stimulus belonged to a different colour category than the rest of the stimuli in the display. The results were clear cut. The participants correctly identified the different stimulus in the between category condition, and they were faster to do so only when it was presented in the right visual field, perceived by the language-dominant left hemisphere. A subsequent verbal interference task abolished the effect, establishing the verbal mediation by the left hemisphere of the colours presented in the right visual field.

Subsequent studies explored the flexibility of the lateralized Whorf phenomenon. Drivonikou, Clifford, Franklin, Özgen, and Davies (2011) taught participants new colour terms for an arbitrary division of the colour space, a separate yellow-green category, and a separate blue-green category. Participants who received this training showed faster response times for the newly trained categories relative to a control group that received no training, but this advantage only appeared when stimuli were presented in the right visual field.

Franklin and colleagues tested the hypothesis that all information is verbally mediated by looking at the phenomenon developmentally. A prevalent research question in this area is where the colour categories employed in language come from. One view holds that natural categories like colour are innate and universal (Franklin, Clifford, Williamson, & Davies, 2005). The alternative view holds that colour categories are language-derived and vary from culture to culture (Roberson, Davidoff, Davies, & Shapiro, 2004). According to this line of investigation, colour categories are emergent rather than innate conceptual properties, entirely dependent on the ambient language and cultural group the infant is exposed to. Franklin and colleagues (2008) used eye tracking to reveal the dynamic activation of linguistic categories during visual perception. They measured how quickly infants began to look at the different colour target in each colour display. The researchers also gave a naming and a comprehension task to the toddlers, specifically measuring their usage and understanding of colour terms in their native language. Based on these data, the researchers split their sample into those children who had acquired the colour terms of their native language, called namers, indicated by high accuracy and frequent usage of colour terms. The other group called learners consisted of children who were still learning the correct understanding and application of their native colour terms. The researchers compared eye-tracking patterns in the two groups. The results showed that those children that could not yet name the colour terms of their native language showed a reverse lateralized Whorfian effect, showing faster initiation time for the left visual field perceived by the right hemisphere. This is not an entirely expected finding; one would have expected no effect, but it may point to a pre-linguistic, innate categorical division of the colour spectrum. However, if such an innate set of categories exists, it is later completely substituted by a language derived set of categories that reside in the language-dominant left hemisphere. This is shown by the eye gaze patterns of the group of namers, which revealed faster initiation time for stimuli presented in the right visual field, perceived by the language dominant left hemisphere. So, the evidence shows that there may be an innate predisposition to perceive colour categories via the right hemisphere. However, once we acquire colour terms, it seems that language influences the functional organisation of colour categorical perception in the brain.

A developmental question arising from this research asks what happens to the pre-linguistic right hemisphere (RH) categories, assuming that they were there in the first place? If replaced, then language is necessary for categorization. If suppressed, then they should (re)surface when language is not available. Paluy, Gilbert, Baldo, Dronkers, and Ivry (2011) hypothesized that aphasias resulting from lesions affecting the language areas of the brain would have an analogous effect to verbal interference in normal populations. The researchers administered a Lateralized Whorf colour experiment to 15 aphasia patients. The patients showed a reversal of the Lateralized Whorf effect: Instead of an advantage for between-category targets in the right visual hemisphere (RVF; processed by the LH), the left visual hemisphere (LVH; processed by the RH) showed an advantage for between-category discriminations.

Explaining the LVF (RH) advantage for between-category discriminations in patients with aphasia is not straightforward. Is this utilization of pre-linguistic RH categorization (as shown in infants)? Or, is this reorganization of brain function following trauma, in which the undamaged hemisphere assumes some of the functions of the damaged hemisphere? A tentative answer comes from corpus callosotomy, that is, severing the corpus callosum that connects the two hemispheres to relieve patients from cases of severe epilepsy. Callosotomy keeps seizure activity from spreading, but it also eliminates direct communication between the two hemispheres. Patient JW underwent a two-stage callosotomy operation in 1979–1980 for epilepsy. There was no clinically observable neglect or aphasia, and JW had little difficulty understanding and performing Lateralized Whorf tasks (Gilbert and colleagues, 2006). JW was faster for between- than within-category discriminations in the RVF (LH), with no difference in the LVF (RH), confirming the lateralized Whorf prediction pattern in normal controls. If there were any pre-linguistic categories in the RH still present then we would have expected to see a between-category advantage in the LVF as well. Such findings have been corroborated with stimuli beyond colour. Patient VP underwent a two-stage callosotomy for the control of epilepsy at the age of 27 years. Her post-surgery intelligence fell within a normal range. This patient was tested in a Lateralized Whorf experiment using cat-dog stimuli (Gilbert, Regier, Kay, & Ivry, 2008). Like patient JW, patient VP was faster for between- than within-category discriminations in the RVF (LH), with no difference in the LVF (RH). Taken together, the findings of the absence of a categorical perception effect in the RH in callosotomy patients are suggestive of the possibility that the RH categorical perception effect in patients with aphasia may come about as a result of functional reorganization of linguistic categorization in the RH following LH brain damage, rather than any language-independent pre-existing categories. Clearly, further evidence from studying categorical perception in language pathology is needed to establish this possibility empirically.

The above evidence notwithstanding, it should be pointed out that not all studies have found evidence for a lateralized category effect for colour. In what is now considered a seminal replication study, Witzel and Gegenfurtner (2011) consistently found categorical perception bilaterally, casting doubt on the purported linguistic basis of the phenomenon. In versions of the lateralized Whorf experiment using the original renderings of the colour stimuli used in Gilbert and colleagues (2006) and in Drivonikou and colleagues (2007), and in versions carefully controlling for lightness and saturation by keeping them constant, they failed to obtain a single lateralized Whorf effect (i.e. categorical perception exclusively in the RVF). They attributed the findings of the lateralized Whorf studies to poor reporting and controlling of the physical characteristics of the colour stimuli, although it is curious that such less than rigorous control would in the vast majority of studies yield RVF dominant categorical perception and not for example LVF dominant or bilateral.

In an earlier study, Roberson, Pak and Hanley (2008) had unravelled the temporal dynamics of bilateral colour categorical perception by splitting their participant groups into fast and slow responders. The results revealed that the former group showed a lateralized effect in the RVF exclusively, whereas the latter group showed categorical perception in both visual fields. The authors interpreted this finding as showing cross-hemisphere transfer via the corpus callosum, in the slow respondents, allowing for the top down information from language to spread contra-laterally over time. Witzel and Genenfurtner (2011) acknowledged Roberson and colleagues' study, but dismissed the possibility of cross-callosal transfer in their own dataset on two premises.

Firstly, although the authors did not split their group into fast and slow respondents, as Roberson and colleagues did, they correlated average reaction times with the lateralized category effect. The lack of a significant negative correlation was interpreted by Witzel and Genenfurtner (2011) as providing null evidence for an emergent bilateral effect as reaction times increase. However, using the lateralized category effect as an index of cross-callosal transfer leaves the issue open, because the variable in question is calculated by subtracting the categorical perception effect (itself calculated by subtracting RTs for across category stimuli from RTs for within category stimuli – the larger the number the stronger the categorical distinction) for the LVF from the categorical perception effect for the RVF. A positive score indicates lateralized categorical perception to the RVF (the classic lateralized Whorf effect), a negative score indicates lateralized categorical perception to the LVF, and a score of 0 (or close) indicates bilateral distribution (i.e. categorical perception in both the LVF and the RVF). However, correlating this variable with reaction times fails to capture the temporal dynamics of cross-hemisphere transfer, because there is no reason to expect that the categorical perception effect in the RVF will be correlated with RTs. The expectation, rather, is that if cross-callosal transfer occurs, then this will be indexed in the LVF scores, such that the longer the RTs, the greater the categorical perception effect will be in the LVF (accounting for the time it takes for cross-callosal transfer to occur). Indeed, when Roberson and colleagues correlated the categorical perception effect in the LVF and overall response time in their Korean speakers, the correlation was significant and strong ($r = .55$), indicating that categorical perception in the LVF was associated with slower reaction times.

Secondly, Witzel and Gegenfurtner (2011) reported average RTs for the implementations of Gilbert and colleagues (2006) that were much higher than that study, but dismissed the possibility that the higher RTs could be indicative of cross-callosal transfer on the basis that the RTs reported in their paper are faster than the RTs reported in Roberson and colleagues (2008). However, Witzel and Genenfurtner's (2011) study was a replication of

Gilbert and colleagues (2006), not Roberson and colleagues (2008). Given the importance the authors place on precise physical rendering of stimuli across studies, we cannot rule out cross-callosal transfer as a basis for the bilateral Whorfian effect they report.

It would be interesting to see the more informative analyses performed, both in terms of splitting the group into slow/fast respondents, and in terms of correlating LVF categorical perception with RTs. However, even if these analyses show no evidence for cross-callosal transfer, they would still not address a basic tenet of the Whorfian hypothesis that all lateralized Whorf studies using a single population of language speakers fail to address; namely differences in speakers of different languages that are isomorphic to cross-linguistic differences in terminology. Even with imperfectly calibrated stimuli, why is it that Greek and English speakers in Thierry and colleagues (2009) and Greek and German speakers in Maier and Abdel Rahman (2018) perceive differently the exact same light and dark blue stimuli, consistent with the colour terminology in their native languages? And that it is even possible to see changes in this sensitivity with increasing levels of bilingualism (Athanasopoulos and colleagues, 2010; see next section for a review of these three studies)? Of more relevance to the lateralized Whorf effect, why is it that the Korean speakers in Roberson and colleagues (2008) perceive differently the exact same green and yellow-green stimuli (consistent with the lexical boundary in Korean) than their English-speaking counterparts? Even if stimuli are not as well controlled as they could be, Whorfian effects across populations still present, which would be difficult to explain by any account outside of the observed cross-linguistic differences between the languages that the populations under study use.

Nonetheless, the mechanism underpinning the purported lateralized Whorf effect is not yet fully understood, with some scholars even questioning the idea that left lateralization of categorical perception is necessarily linked specifically to linguistic processing, even when such lateralization is found (Holmes & Wolff, 2012). A meta-analysis approach may indicate the pervasiveness of this phenomenon across different domains and stimuli, and hopefully elucidate its neural basis further.

2. Neuroscientific Approaches to Linguistic Relativity: neural measures

2.1. ERP evidence

Event-related potentials (ERPs) are measures of particular interest for linguistic relativity research since they provide an online measure of the different stages of visual and/or auditory processing without the need for overt responses. ERPs thus allow us to track the temporal unfolding of automatic and unconscious cognitive processes in response to particular events (e.g., presentation of words, pictures, sounds). ERPs are measured by averaging the electrical activity generated by the brain (i.e., electroencephalographic signal) in response to time-locked external stimuli. ERP waveforms elicited by visual or auditory stimuli can be sub-divided into different components underlying specific cognitive processes represented by positive and negative deflections with specific temporal and topographical distribution. ERPs' particular sensitivity to the temporal unfolding of the different stages of cognitive processing allows for the differentiation between effects triggered at early perceptual stages and effects triggered later in the processing stream, when the online activation of linguistic labels might affect stimulus processing.

Categorical perception mediated by linguistic labels is one of the most studied phenomena using ERPs. This is widely used in combination with the visual oddball paradigm, which provides a measure of an automatic response of the brain to stimuli that deviate from a sequence of frequent stimuli. The waveform deflections generated by the deviants compared to standard stimuli elicit the deviant-related negativity (DRN). DRN typically occurs between 100 and 350 ms post-stimulus and can be decomposed into different sub-components corresponding to an early perceptual stage of object identity resolution peaking around 150 ms and a later perceptual processing stage modulated by attention allocation and cognitive control with an onset ~200 ms (Folstein & Van Petten, 2008). When the stimuli used in this paradigm are presented in the peripheral visual field to a centrally presented task, this deviant-related negativity is usually referred to as the visual mismatch negativity (vMMN, Czigler, 2014). This ERP component is thought to index automatic and pre-attentive processing of a deviant stimulus compared to the memory trace of the preceding standard stimuli (Kimura, Schröger, Czigler, & Ohira, 2010) and it peaks around 140 ms in the form of a negative deflection at posterior electrode sites. It is worth noting that the vMMN component is typically elicited before lexical access takes place, which typically occurs after 200 milliseconds of picture onset

(Blackford, Holcomb, Grainger, & Kuperberg, 2012; Costa, Strijkers, Martin, & Thierry, 2009), or of word onset (Grainger & Holcomb, 2009).

One of the effects of language on perception that has been studied using ERPs is that of perceptual colour categorization (see Thierry, 2016, for a review). Mo, Xu, Kay, and Tan (2011) investigated the lateralized Whorf hypothesis using ERPs in combination with the oddball paradigm. In this study participants were asked to perform a visual detection task -i.e. detection of target circles appearing in the middle of the screen, 10 % of trials, amongst consecutive fixations crosses-, whilst coloured squares appeared peripherally in the left and right visuals fields simultaneously (LVF and RVF). As deviant stimuli, the authors selected four colours of two Chinese colour categories (green/blue) of constant saturation and brightness (neutral green, bluish green, greenish blue, and neutral blue). Deviants (20% of the trials) were presented across four blocks paired with standard stimuli (70% of the trials) and belonged either to the same or a different colour category (green or blue). Crucially, deviant stimuli were presented 10% of the times in the RVF and 10% of the times in the LVF. In line with previous behavioural evidence (see Gilbert and colleagues, 2006), results showed greater vMMN for between category than within category colour pairs presented in the RVF, but not in the LVF. This thus provides further evidence that perceptual categorization of colour is predominantly processed by the left hemisphere, where most language functions are hosted. Mo and colleagues' (2011) study provides the first electroencephalographic evidence for the lateralized Whorfian effect, suggesting that the effects of language on perceptual categorization of colour are automatic, occur early in the visual processing stream, and are pre-attentive in essence.

Xia, Xu, and Mo (2019) following a similar design, further investigated the left-lateralized Whorfian effect by testing a different language system, sign language. Sign language comprehension heavily relies on visual spatial processing and this leads to more bilateral activation when processing linguistic stimuli (MacSweeney and colleagues, 2002; Newman, Supalla, Hauser, Newport, & Bavelier, 2010). Hence, the authors hypothesized that categorical perception of colours for native signers should recruit language-linked regions from both the left and the right hemispheres. They tested 14 severely-to-profoundly deaf native signers performing a similar task as Mo and colleagues (2011) whilst ERPs were recorded. As predicted, results showed greater vMMN amplitudes for between- than within-category colour pairs and these were consistently present in both the LVF and the RVF. This thus provides the first theoretically driven bi-lateral Whorfian effect and demonstrates that the language (or language system) that we use to communicate has a direct impact on categorical perception.

It is worth noting that previous electrophysiological studies had already shown perceptual effects of language labels during categorical perception of colours (Clifford, Holmes, Davies, & Franklin, 2010; Fonteneau & Davidoff, 2007; Holmes, Franklin, Clifford, & Davies, 2009; Liu and colleagues, 2010; Liu and colleagues, 2009). It was argued that language-defined colour boundaries can influence top-down attention allocation resources to enhance further perceptual processing of relevant stimuli. However, Mo and colleagues' (2011) and Xia and colleagues' (2019) studies provide new insights into the role of language use and organization in categorical perception by demonstrating that the lateralized Whorfian effect can also affect pre-attentive stages of processing (peaking around 160ms in both studies), corroborating earlier studies that had established a pre-attentive effect of language in speakers of different languages, and in bilinguals (more on this later).

Along the same line, Clifford and colleagues (2010) and Forder, He, and Franklin (2017) showed that lexical labels affect categorical perception of colours at two different stages, an early pre-attentive stage and a later stage of processing. Both of these studies found greater deviancy effects for between- than within-category discrimination based on participants' self-perceived colour boundaries starting as soon as 100 ms, suggesting that categorical relationships between colours are traceable to early sensory stages of visual processing. This is then followed by later deviancy effects where lexical retrieval of the corresponding colour labels might take place, re-enforcing categorical distinctions.

Although the studies mentioned above provide evidence for colour categories affecting pre-attentional and perceptual processes, the most compelling evidence for Whorf's hypothesis should derive from cross-linguistics studies. As stated above (see section 1.3), different languages can hold labels representing different categorical boundaries. Hence, speakers of different languages should present pre-attentive categorical perception in line with their corresponding linguistic distinctions. The first ERP study investigating cross-linguistic influences in categorical perception of colour is that of Thierry, Athanasopoulos, Wiggett, Dering, and Kuipers (2009). In that study, the authors presented to Greek-English bilinguals and English monolingual participants a stream of circles and squares that could either be light blue, dark blue, light green, and dark green. Participants had to press a button key every time a square was presented on the screen independently of the colour. These target trials were disregarded from the analyses. 10% of the critical trials (circles) had a change in luminance (dark vs light) compared with the standard critical trials (70%). Greeks participants, who have two terms (*galazhio* and *ble*) to differentiate dark and light blue but only one term for dark and light green (*prasino*), showed greater vMMN modulations for blue deviants than for green deviants. Whilst English native speakers did not statistically differ in their vMMN amplitudes for the two different types of colours (the language group x colour x deviancy interaction was significant). This was taken as evidence that lexical labels lead to greater perceptual discrimination of colours in an automatic and unconscious manner.

Furthermore, the same group of researchers (Athanasopoulos, Dering, Wiggett, Kuipers, & Thierry, 2010) re-analysed the data based on participants' exposure to English language. They split the group of Greek-English bilinguals in two based on the length of stay in the United Kingdom. Results showed that Greek speakers who had been more exposed to English language environments (18 months or longer) had reduced deviancy effects, mirroring those of English monolingual participants. This study not only showed that language affects perceptual processes cross-linguistically, but that acquiring a new language shifts existing categorical colour boundaries affecting early perceptual discrimination processes. This study also corroborated that categorical perception of colours is a dynamic process that can be acquired and re-shaped during the lifespan based on language experience.

In this regard, more recent studies have shown that categorical perceptual effects can also be found with newly trained colour categories. For instance, Zhong, Li, Li, Xu, and Mo (2015) trained Chinese speakers to associate two blues (B1 and B2 previously used in Gilbert et al., 2006) to two different lexical labels see also (see also Zhou and colleagues., 2010). After extensive training participants were asked to perform a visual detection task similar to that of Mo and colleagues (2011). Deviants from this study were formed by four new blues from the same two novel categories (B11/B12 and B21/B22). Results showed greater vMMN effects for between-category than within-category stimuli in the RVF but not in the LVF for newly acquired categorical boundaries. Crucially no deviancy effects in either visual field were found for a control group of Chinese speakers who did not undertake any training. Hence Zhong and colleagues (2015) suggested that recently acquired colour terms via training produce similar lateralized Whorfian effects to those acquired in natural language. This study thus provides further evidence that lexical categories enhance perceptual sensitivity to colour boundaries.

Moreover, it suggests that the links between language and categorical perception of colours are highly plastic in nature and can be re-shaped within short periods of time.

More recently, Maier and Abdel Rahman (2018) used the same blue and green stimuli that were used in Thierry and colleagues (2009), to investigate whether linguistic terminology may enhance visual consciousness. In a standard iteration of the attentional blink paradigm (in which targets are often not consciously noticed), participants had to successively report the direction of a semi-circle (coloured grey, always present) and the direction of a triangle (coloured light/dark blue, absent in 18.2% of the trials, but always presented after the semi-circle when present) occurring in a rapid serial visual presentation stream of distractors (grey polygons other than semi-circles or triangles) presented against a light/dark blue background (light/dark green contrasts were also used as control stimuli, as in Thierry and colleagues, 2009). ERPs were time-locked to the onset of the dark/light blue triangle, averaged across time windows of interest. Results showed that in the P1 (early visual processing component), mean amplitude was larger in the blue compared to the green conditions, consistent with the lexical contrast in blues but not greens in Greek. German speakers, whose language only has one basic term for blue, did not show such a perceptual advantage for blues over greens. There was also an amplitude difference between blues and greens in the N2 ERP component, indexing visual consciousness, exclusively in Greek speakers. The authors concluded that language plays an important role in what we consciously perceive.

ERP evidence has also been brought to bear on the debate regarding whether colour categories are language independent or not (see previous section). For instance, Clifford, Franklin, Davies, and Holmes (2009) presented schematic coloured faces to 7 month old infants whilst ERPs were recorded. Three types of colours varying in hue from green to blue were presented in an oddball paradigm, so that deviants could either belong to the same (green) or different (blue) colour category than the standards (green). Results showed deviancy effects for between-category distinctions but not for within-category in the form of Nc (i.e., Negative-central wave-forms peaking between 250-650ms) and NSW/PSW (i.e., negative and positive slow wave peaking between 1150-1700ms) components. These results suggest that different-category deviants elicited greater attention allocation and triggered the detection of novelty in pre-verbal infants. Based on these results Clifford and colleagues (2009) suggested that colour categorization can occur without language and that colour categories in language may partially originate from a common universal constraint on colour categorization (see also Kay & Regier, 2003). More recently Yang, Kanazawa, Yamaguchi, and Kuriki (2016) using near-infrared spectroscopy (NIRS) found similar colour category effects with pre-verbal infants (5 to 7 months old) using a different paradigm. In this study, infants were presented with coloured geometrical figures that changed colour every 1 second for a period of 10 seconds during each trial. Within each trial, colour alternated at 1 Hz from within the green or blue categories (e.g., G1/G2) or between them (e.g., G1/B1). As in Clifford and colleagues (2009), results showed categorical perceptual effects for the between-category condition, but not for the within-category. They found that alternations of colours between colour categories resulted in increased concentration of oxyhemoglobine and deoxy-haemoglobin in bilateral occipito-temporal regions compared to baseline conditions. Interestingly, the authors replicated the same results with a population of young adults. This led Yang and colleagues (2016) to conclude that colour categories may develop before language acquisition and therefore they exist independently of language. However, although the studies by Clifford and colleagues (2009) and Yang and colleagues (2016) with pre-verbal infants could help us to better understand the origin and development of colour categorical perception, the current studies present a series of limitations. Critically, they fail to dissociate the mechanisms underlying categorical perception of colours from those of language exposure to categorical boundaries of the native language and the inherent statistical probabilities of the environment (e.g., parents' directed speech emphasizing colour boundaries). It is well established that pre-verbal infants acquire sophisticated information about the properties of language by simply listening to language (see Khul, 2000, for a review), and that language experience warps categorical perception of speech sounds very early in development. For instance, at 6 months, infants are able to detect phonetic prototypes from non-prototypes of the same category that are specific to the native language (Kuhl, 2000), and at 8 months infants can segment words from fluent speech (Saffran, Aslin, & Newport, 1996), and engage in cortical tracking of temporal information in directed speech which facilitates speech encoding (Kalashnikova, Peter, Di Liberto, Lalor, & Burnham, 2018). More importantly, Bergelson and Swingley (2012) have shown that native language learning in pre-verbal infants goes beyond the acquisition of the sound structure. At 6 months, even before infants begin to babble, they have already started to link words with their referents (e.g., food, body-parts), suggesting that the learning of conceptual and linguistic categories might go hand in hand, starting in initial stages of development. Hence, one could argue that effects of categorical perception of colours in pre-verbal infants should be taken with caution since they may coincide with mapping of linguistic categories already present very early on in development. We suggest that future research investigating categorical perception of colour in pre-verbal infants should explore colour categories not present in the native language.

A related caveat concerns the provenance of the colour stimuli used for perceptual discrimination in infant studies. A basic assumption of categorical colour perception is that one pair of colors that straddles a category boundary is discriminated better than a *psychophysically equidistant* pair of colours that is on one side of the boundary. It has recenty been pointed out that this assumption of psychophysical equidistance may be compromised (Oceláк, 2016; Forder & Lupyan, 2019). Specifically, the colour space used in infant studies is calibrated on adults. If language has already warped adult perception and these biases become part of the space, then the categorical perception that infants show simply reflects sensitivity to the adult categories, rather than revealing any pre-existing perceptual differences.

In line with the electrophysiological evidence showing that language labels affect categorical perception of colours, Boutonnet, Dering, Vinas-Guasch, and Thierry (2013) investigated the neural correlates of object categorisation using the oddball paradigm. As stated previously, different languages differ in the way they categorise objects, and these distinctions highlight the characteristics of the objects themselves affecting categorization and decision-making processes (e.g. Ameel, Malt, Storms, & Van Assche, 2009; Pavlenko & Malt, 2011). In the study by Boutonnet and colleagues (2013), Spanish and English speakers performed an object detection task with pictures depicting mugs and cups. It is worth noting that Spanish has only one word for both mug and cup, “taza”. Participants were asked to detect the odd stimulus (a bowl) which appeared 15% of the times from a string of cups (standards: 70% of the stimuli) and mugs (deviants: 15% of the stimuli). Participants performed two blocks of trials so that both cups and mugs were presented once as standards and once as deviants. As expected, results showed an early deviancy effect only for English participants whilst no significant deviancy effects were observed for the Spanish group. The authors of this study suggested that language is functionally integrated with other perceptual systems in the human brain. Following these premises, it is not surprising that once language is acquired language-specific terminology affects the way objects are perceived and categorised.

More recently, Boutonnet and Lupyan (2015) also investigated whether the perception of familiar animals and artefacts could be enhanced by priming the images with spoken words (e.g., the picture of a guitar preceded by the word guitar) or equally familiar nonverbal sounds (e.g., the picture of a guitar preceded by the sound of a guitar). The authors predicted that since the relationship between words and their conceptual representations is categorical, hearing the labels should facilitate object recognition at early perceptual stages. As predicted, images preceded by labels elicited more positive P1 peak amplitudes, an index of low-level visual features processing, than when they were preceded by non-verbal sounds. The authors interpreted these results as evidence for a label advantage on the grounds that labels act as category-level priors that alter the processing of subsequent visual information. However, early visual components such as the P1 are generally known to be sensitive to stimulus expectation, independently of the categorical relationship between stimulus and context (e.g., prime). For instance, Federmeier and Kutas (2001) showed anterior modulations of the N1 (the anterior counterpart of the posterior P1) for exemplars expected on the basis of contextual information (in this case, high cloze probability sentences) as compared to unexpected exemplars, independently of whether the unexpected exemplar belonged to the same or a different category. For instance, the word ‘guitar’ arguably provides more contextual information than the sound of a guitar, a specific characteristic associated with the object (i.e., a musical sound), leading to a different priming effect. Alternatively, word-picture association could be stronger than single attribute-to-picture association due to frequency of use and familiarity of words compared to sounds. It is thus unclear whether the differences found relate to a frequency effect or a label advantage per se. It is worth noting that in the Boutonnet and Lupyan (2015) study, spoken words and non-verbal cues from the same category did not elicit differential N400 modulations. This thus suggests a similar degree of target expectations across the two types of cues.

Relatedly, Maier, Glage, Hohlfeld, and Abdel Rahman (2014) demonstrated that recently learned verbal category boundaries of unfamiliar objects can also influence early stages of perceptual processing (see also Rabovsky, Sommer, & Abdel Rahman, 2012). In their study participants were asked to learn the labels of two pairs of similar-looking made-up objects, two within and two between category. Furthermore, in order to investigate the role of semantically enriched verbal labels in categorical perception, half of the objects were further provided with semantic knowledge. Two to three days after learning participants performed a lateralised oddball paradigm whilst brain potentials were recorded. The results of this study showed greater P1 (the first positive going ERP component related to visual processing) modulations for objects belonging to different category than within category in the RVF, but not in the LVF, and this was consistent for label only and semantically enriched conditions. This study thus provides further evidence of the online influence of language on perception by showing that recently acquired verbal category boundaries affect low-level perceptual processing.

Similarly, Yu, Mo, Zeng, Zhao, and Mo (2017) trained Chinese participants to associate novel shapes (i.e., irregular polygons) with semantically irrelevant Chinese monosyllables. After training participants were asked to perform a lateralised oddball task whilst ERPs were recorded. Results of this study also showed greater vMMN for different-category deviants presented in the RVF compared to same-category deviants, whilst no difference was observed for different vs same category deviants presented in the LVF. Hence Yu and colleagues (2017) provided further evidence for the lateralized Whorfian effects beyond colour perception, suggesting that language terminology may shape low-level perceptual processes even for short-term trained lexical categories.

3. ERP evidence for grammatical effects on perceptual categorization

Despite the extensive behavioural evidence showing categorical effects originating in language-specific grammatical constructs, such as grammatical gender (e.g. Lemhöfer, Spalek, & Schriefers, 2008) and grammatical aspect (e.g. Athanasopoulos and colleagues, 2015), only few electrophysiological studies have explored perceptual effects beyond those originating in lexical labels. The effects of grammatical constructs on perception is perhaps the most essential question regarding Whorf's original thoughts (see Lucy, 1997). Hence evidence coming from neurophysiological methods which are less susceptible to the involvement of covert language processes, such us conscious verbal strategies, are crucial.

One attempt to show grammatical effects on categorical perception of objects using ERPs is that of Boutonnet, Athanasopoulos, and Thierry (2012). In that study, the authors investigated the neural substrates of unconscious grammatical gender processing during object categorization. Spanish-English bilinguals and control English monolinguals were presented with triplets of pictures depicting everyday objects. Participants were asked to perform a semantic categorization task based on whether the third picture belong to the same or different category from the previous two pictures (e.g., Tomato-Celery-Asparagus). Unbeknownst to the participants the target picture was either consistent or inconsistent with the grammatical gender of the previous two objects when translated into Spanish. Although behavioural priming effects were not found, both groups of participants showed, as expected, greater N400 modulations, which indexes semantic integration (Kutas & Federmeier, 2011) for semantically unrelated targets than for semantically related ones. Critically, incongruent grammatical gender targets elicited greater modulation of the Left-Anterior Negativity (LAN), an ERP marker of morphosyntactic processing (Friederici & Jacobsen, 1999; Friederici, Pfeifer, & Hahne, 1993), only for Spanish-English bilinguals. The language context of the study was all-in-English and hence Spanish grammatical gender information was not needed to perform the task. Nonetheless Spanish-English bilinguals showed automatic and unconscious retrieval of non-relevant grammatical categories. Boutonnet and colleagues (2012) thus demonstrated that speakers of gendered languages may unconsciously retrieve entrenched information related to grammatical categories such as gender when perceiving and extracting semantic information of everyday objects.

Even more compelling evidence for the effects of grammatical constructs on perceptual processes is provided in the study by Flecken, Athanasopoulos, Kuipers, and Thierry (2015), where the authors generalised the effects of language on perception to the dynamic domain, i.e., motion events. In this study, native English and German speakers were asked to perform a matching task whilst ERPs were recorded. On each trial, participants saw an animated motion event followed by a picture symbolizing a match trajectory and end-point (5% of the trials), a mismatch trajectory and end-point (75% of the trials), a match on trajectory only (10% of the trials) and a match on end-point only (10% of the trials). Participants were asked to respond only to the pictures depicting full match conditions. German, differently than English, is a non-aspect language (see section 1). Following the premise that motion event verbalization and cognition is more goal-oriented in speakers of non-aspect languages than in speakers of aspect languages (see section 1), Flecken and colleagues (2015) predicted that German speakers should perceive end-point matched pictures more saliently than trajectory only. The results were clear cut: English participants showed no differential P3 amplitudes for the critical conditions, whereas German speakers showed greater P3 amplitudes in the end-point match than the trajectory match conditions. Authors suggested that these results show an attentional bias during the perception of motion events towards aspects highlighted and entrained by language.

3.1. Structural and functional MRI evidence

Despite increased neuroscientific evidence for links between language and thought, studies directly investigating linguistic relativity employing structural or functional magnetic resonance imaging (MRI) methods are still scarce. MRI provides anatomical information about the structure of the brain whilst functional MRI (fMRI) provides indirect measures of neural activity based on the level of oxygenation of haemoglobin in the brain. When neurons are more active they require oxygenated blood and the resulting change of oxygen concentration can be detected by the scanner. Tan and colleagues' (2008) pioneering fMRI study investigated whether the areas of the brain implicated in language processing such as left posterior temporo-parietal regions (Duffau, 2008; Edwards and colleagues, 2010) modulate the neural systems mediating the perceptual discrimination of colours (i.e., visual occipital cortex; Davidoff & Ostergaard, 1984; Oxbury, Oxbury, & Humphrey, 1969; Rentzeperis, Nikolaev, Kiper, & van Leeuwen, 2014). In Tan and colleagues (2008), Chinese speakers were asked to perform a colour discrimination task involving easy-to-name and hard-to-name colours that appeared in pairs on the screen for 100ms followed by a mask with a duration of 900ms. Participants were asked to respond by pressing a button, and thus no naming was required during this task. The findings showed that visual occipital regions involved in perceptual processing of colour, as well as brain regions involved in language processing were co-activated. Moreover, perceptual discrimination of easy-to-name colours elicited stronger activation in regions responsible for word-finding processes, namely the left-superior temporal gyrus and left-inferior parietal cortex (Desmurget and colleagues, 2009; Sonty and colleagues, 2003). Based on those findings Tan and colleagues (2007) suggested that perceptual discrimination of colours involves co-activation of bi-lateral visual cortices and left-lateralized posterior temporo-parietal regions implicated in object naming in an automatic manner, even when colour label retrieval is not necessary, that is during perceptual decisions (see also Francken, Kok, Hagoort, & de Lange, 2015, for similar fMRI evidence showing language regions involvement during perceptual decisions of motion events).

Based on the left-lateralised language regions' involvement during perceptual discrimination of colours reported by Tan and colleagues (2008), Siok and colleagues (2009) went one step further by directly investigating the lateralised Whorfian effect found in previous behavioural and ERPs studies using event-related fMRI. To do so, Siok and colleagues (2009) used a modified version of Gilbert and colleagues' (2006) visual search task in Chinese participants. Behaviourally, this study showed stronger categorical perception in the RVF than in the LVF, consistent with studies using similar designs (Drivonikou and colleagues, 2007; Gilbert and colleagues, 2006), although the key interaction only approached significance.. More interestingly, fMRI results showed greater activation of language-related areas (left-lateralised posterior temporal and inferior prefrontal regions, and middle temporal gyrus) for between-lexical-category than for within-lexical-category targets, but only when they were presented in the RVF. Furthermore, when the different-lexical-category targets were presented in the RVF, Siok and colleagues (2009) also observed greater activation of visual colour regions (V2/3), coinciding with increased activity of the posterior temporo-parietal region. The authors thus suggested that language influences the discrimination of colours by enhancing the activation levels of the visual cortex implicated in colour perception.

In line with fMRI evidence that categorical perception of colour provokes orchestrated cortical activity involving language regions and visual cortices, Kwok and colleagues (2011) recently demonstrated that these neural circuits can be modified by learning new colour categories. In their study participants learned newly defined subcategories of the categories green and blue in a period of 2h. Results showed increased volume of grey matter in V2/3 of the left visual cortex, the same region shown by Siok and colleagues (2009) to mediate in colour vision. Although the results from Kwok and colleagues (2011) should be taken with caution (e.g., there was no control group performing the training without language labels), their finding corroborates that the anatomical structure of the adult human brain is highly plastic (see Wenger and colleagues, 2017) and can change very quickly, specifically during the acquisition of new colour category boundaries. It corroborates behavioural and ERP findings of the effects of short-term training on perceptual discrimination and categorisation, and similar long-term effects in bilinguals, reviewed earlier in this paper.

It would appear then, that categorical perception can be shifted by acquiring new language labels that in turn will modify the underlying cerebral structures. Other studies additionally show that the brain can also be stimulated or impaired by using techniques such as transcranial direct current stimulation. For instance, following the evidence provided by the previous neuro-anatomical and neuro-functional studies, Lupyan, Mirman, Hamilton, and Thompson-Schill (2012) showed that categorical perception of existing categorical labels can be temporally impoverished by applying direct current stimulation (cathodal) to the left inferior frontal cortex. Cathodal stimulation is thought to decrease excitability of the sites of interest (e.g., Broca's area), whereas anodal stimulation is thought to increase it (Nitsche & Paulus, 2000). Similarly, Perry and Lupyan (2017) showed that cathodal stimulation of Wernicke's area, a region of the brain (left-lateralized superior temporal lobe) implicated in language comprehension, slightly reduced accuracy of sparse categories in a picture-word verification task, whilst increasing overall speed of response regardless of category type (sparse or dense). Furthermore, anodal stimulation over the same area decreased the verification speed to trials of sparse categories (see also Perry & Lupyan, 2014).

Summary and Conclusions

The behavioural and neuroscientific findings summarized and discussed in this paper are in accordance with recent neuroscientific evidence outside of the Whorfian paradigm (e.g. Pulvermüller, 2018; but see also Mahon & Hickok, 2016), converging on a view of information processing that is characterized by massive interactivity inherent in the functional organisation of the human brain. It is also in line with accounts of general top-down modulation of perceptual processes found in many other knowledge representations (beyond language, see e.g. Lupyan, 2015 for a review). The studies reviewed here show that language affects essentially all stages of the visual processing stream. From later stages of visual integration and semantic interpretation, where the effect of language would be expected, given the rich behavioural evidence, to earlier stages of visual perception, where brain signatures traditionally thought of as impervious to top-down modulation (vMMN, N1) are in fact shown to be susceptible to such modulation by the observer's language background, in predictable ways.

A criticism often levelled at studies showing evidence of language effects on categorical perception and other visual discrimination behaviour is that such effects are 'trivial' 'weak', or even 'banal', because underlying biological structures of vision (and other relevant senses) remain unchanged. But such value judgments would only be true if indeed the Whorfian question predicted structural perceptual changes at a biological level. One consequence of adopting such an interpretation of the Whorfian thesis is that it easily gives rise to strawman argumentation, where the Whorfian hypothesis is recast as some kind of radical, even ridiculous assertion that language rewires the biological physiology of the vision system, and then easily dismissed. This quote from Pinker is a case in point:

"No matter how influential language might be, it would seem preposterous to a physiologist that it could reach down into the retina and rewire the ganglion cells." (Pinker, 1995, p. 62).

However, no-one is claiming that this ‘preposterous’ thesis could ever be true. Whorf never made any claims about the physiology of the human eye, or even the structure of the brain. Rather, the Whorfian question, in our view, concerns language-driven human behaviour. The endeavour, from a neuroscientific point of view, is to unravel the neural correlates of such linguistically-driven behaviour, thus revealing its biological basis. Dismissing or accepting the Whorfian hypothesis solely on the basis of whether we can find effects on brain physiology or not is a garden path endeavour, because all behaviour (including language-driven behaviour) ultimately has a physiological basis, even if not captured (yet) by the available instruments and techniques used. Even in studies that claim to report anatomical changes (e.g. Kwok and colleagues, 2011), interpreting the cause and direction of the anatomical change is not straightforward, and such findings seem to only tentatively demonstrate that language induces anatomical changes in the brain.

Be that as it may, the evidence suggests that language does have profound effects on the physiology of the brain in terms of neural plasticity (e.g. Hebb, 1949)). This is reflected behaviourally in any simple perceptual discrimination task and neurophysiologically in fMRI/ERP/etc. studies that reveal the neural correlates of such changes in behaviour as a function of language. Modulation of early components of visual perception are highly indicative of neural adaptation not only as a function of newly learned categories in the laboratory (Zhong and colleagues, 2015), but also as a consequence of long-term exposure and usage of another language, as in the case of bilingualism (Athanasopoulos and colleagues, 2010).

Another consequence of adopting a radical interpretation of the Whorfian hypothesis as a question of biological anatomy is that it makes it possible to classify the multitude of modern evidence of Whorfian effects (the veracity of which cannot otherwise be denied except by rigorous replication) as methodologically superficial . As Pinker (2007) put it:

‘‘speakers of different languages tilt in different directions in a woolly task, rather than having differently structured minds’’ (p.148).

Again, it is far from clear what kind of evidence would constitute a marker of ‘differently structured minds’, apart perhaps from what the author chooses it to be. But if one defines physiological changes in terms of neuroplasticity, and then looks at the first brain signature of unconscious change detection, speakers of different languages show different modulations, in line with their native language's terminology (Thierry and colleagues, 2009). The effect is very robust and replicable across different samples of native speakers and in bilinguals (Mo and colleagues, 2011).

In any case, the empirical evidence from multiple domains of enquiry points to an information flow in the brain that is akin to a continuum: a conveyor belt of incoming information that is constantly being modulated by top-down expectations, which in turn are continuously updated by the new information (Lupyan & Clarke, 2015). So, the real question is not whether language can alter the anatomical and biological structures of perception, but rather where in that continuum top-down information like language kicks in. We now know that visual cortex can index modulations in top down information, so the answer to that would be 'as soon as the first brain signatures of visual perception and attention are registered'. Indeed, the first positive brain signature of visual processing (P1) related to visual cortex and of unconscious change detection (vMMN) related to pre-attentive perception present with very robust and replicable evidence of language modulation with a variety of paradigms (e.g. oddball detection, attentional blink), and across different samples of native speakers and in bilinguals. This interpretation is compatible with parallel activation of several brain regions as soon as the visual input is received, as in the case of colour and odour words.

This is also compatible with the basic tenets of Hebbian learning (McCLelland, 2006). During the process of first and second language learning, new links between visual objects with their labels (or events with their descriptive grammatical structures) will emerge by means of the restructuring and reinforcement (everyday language use) of assemblies of neurons connecting both the conceptual representations with linguistic and somatosensorial systems. It is thus unsurprising to observe experimental studies showing effects of language on perceptual processes. The wonder here would be to show that perceptual processes can operate in isolation from the rest of cognitive processes and that they are completely invariant to life experiences and statistical learning. We conclude that if we take into account the existing neuroscientific evidence, then instead of labelling effects of language on early stages of visual integration as 'surprising' or 'unexpected', we may find ourselves expressing our surprise if such Whorfian effects were not found.

References

Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language, 60*(2), 270-290.
Athanasopoulos, P., & Albright, D. (2016). A Perceptual Learning Approach to the Whorfian Hypothesis: Supervised Classification of Motion. *Language Learning, 66*(3), 666-689.
Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive science, 37*(2), 286-309.
Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., . . . Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological science, 26*(4), 518-526.
Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J.-R., & Thierry, G. (2010). Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition, 116*(3), 437-443. doi:10.1016/j.cognition.2010.05.016
Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience, 4*(10), 829.
Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253-3258.
Blackford, T., Holcomb, P. J., Grainger, J., & Kuperberg, G. R. (2012). A funny thing happened on the way to articulation: N400 attenuation despite behavioral interference in picture naming. *Cognition, 123*(1), 84-99.
Boutonnet, B., Athanasopoulos, P., & Thierry, G. (2012). Unconscious effects of grammatical gender during object categorisation. *Brain Research, 1479*, 72-79. doi:10.1016/j.brainres.2012.08.044
Boutonnet, B., Dering, B., Vinas-Guasch, N., & Thierry, G. (2013). Seeing objects through the language glass. *Journal of Cognitive Neuroscience, 25*(10), 1702-1710. doi:10.1162/jocn_a_00415
Boutonnet, B., & Lupyan, G. (2015). Words Jump-Start Vision: A Label Advantage in Object Recognition. *Journal of Neuroscience, 35*(25), 9329-9335. doi:10.1523/JNEUROSCI.5111-14.2015
Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology, 49*, 454-462.
Bylund, E., Athanasopoulos, P., & Oostendorp, M. (2013). Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics, 51*(5), 929-955.
Chomsky, N. (2000). Linguistics and brain science. *Image, Language, Brain*, 13-28. Retrieved from <Go to ISI>://WOS:000170232200002
Clifford, A., Franklin, A., Davies, I. R., & Holmes, A. (2009). Electrophysiological markers of categorical perception of color in 7-month old infants. *Brain and Cognition, 71*(2), 165-172. doi:10.1016/j.bandc.2009.05.002
Clifford, A., Holmes, A., Davies, I. R., & Franklin, A. (2010). Color categories affect pre-attentive color perception. *Biol Psychol, 85*(2), 275-282. doi:10.1016/j.biopsycho.2010.07.014
Costa, A., Strijkers, K., Martin, C., & Thierry, G. (2009). The time course of word retrieval revealed by event-related brain potentials during overt speech. *Proceedings of the National Academy of Sciences, 106*(50), 21442-21446.
Czigler, I. (2014). Visual mismatch negativity and categorization. *Brain Topography, 27*(4), 590-598. doi:10.1007/s10548-013-0316-8
Davidoff, J., & Roberson, D. (2004). Preserved thematic and impaired taxonomic categorisation: A case study. *Language and Cognitive Processes, 19*(1), 137-174.

Davidoff, J. B., & Ostergaard, A. L. (1984). Colour anomia resulting from weakened short-term colour memory: A case study. *Brain, 107*(2), 415-431.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632-1634.
Desmurget, M., Reilly, K. T., Richard, N., Szathmari, A., Mottolese, C., & Sirigu, A. (2009). Movement intention after parietal cortex stimulation in humans. *Science, 324*(5928), 811-813.
Drivonikou, G., Clifford, A., Franklin, A., Özgen, E., & Davies, I. R. (2011). Category training affects colour discrimination but only in the right visual field. *New directions in colour studies, 112*(3), 251-264.
Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., & Davies, I. R. (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.
Duffau, H. (2008). The anatomo-functional connectivity of language revisited: new insights provided by electrostimulation and tractography. *Neuropsychologia, 46*(4), 927-934.
Edwards, E., Nagarajan, S. S., Dalal, S. S., Canolty, R. T., Kirsch, H. E., Barbaro, N. M., & Knight, R. T. (2010). Spatiotemporal imaging of cortical activation during verb generation and picture naming. *NeuroImage, 50*(1), 291-301.
Ellis, H. C., & Shumate, E. C. (1973). Encoding effects of response belongingness and stimulus meaningfulness on recognition memory of trigram stimuli. *Journal of Experimental Psychology, 98*, 70-78.
Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 202.
Flecken, M., Athanasopoulos, P., Kuipers, J. R., & Thierry, G. (2015). On the road to somewhere: Brain potentials reflect language effects on motion event perception. *Cognition, 141*, 41-51. doi:10.1016/j.cognition.2015.04.006
Fodor, J. A. (1975). *The language of thought* (Vol. 5): Harvard university press.
Fodor, J. A. (2008). *LOT 2: The language of thought revisited*: Oxford University Press on Demand.
Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology, 45*(1), 152-170.
Fonteneau, E., & Davidoff, J. (2007). Neural correlates of colour categories. *Neuroreport, 18*(13), 1323-1327.
Forder, L., He, X., & Franklin, A. (2017). Colour categories are reflected in sensory stages of colour perception when stimulus issues are resolved. *PLoS One, 12*(5), e0178097. doi:10.1371/journal.pone.0178097
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.
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. doi:10.1162/jocn_a_00682
Franklin, A., Clifford, A., Williamson, E., & Davies, I. (2005). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of experimental child psychology, 90*(2), 114-141.
Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences, 105*(47), 18221-18225.

Friederici, A. D., & Jacobsen, T. (1999). Processing grammatical gender during language comprehension. *Journal of psycholinguistic Research, 28*(5), 467-484.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive brain research, 1*(3), 183-192.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition, 83*(1), 49-79.

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.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and Language, 105*(2), 91-98.

Grainger, J., & Holcomb, P. J. (2009). Watching the Word Go by: On the Time-course of Component Processes in Visual Word Recognition. *Language and Linguistics Compass, 3*(1), 128-156. doi:10.1111/j.1749-818X.2008.00121.x

Hebb, D. O. (1949). The organization of behavior; a neuropsycholocigal theory. *A Wiley Book in Clinical Psychology.*, 62-78.

Holmes, A., Franklin, A., Clifford, A., & Davies, I. (2009). Neurophysiological evidence for categorical perception of color. *Brain Cogn, 69*(2), 426-434. doi:10.1016/j.bandc.2008.09.003

Holmes, K. J., & Wolff, P. (2012). Does categorical perception in the left hemisphere depend on language? *Journal of Experimental Psychology: General*, *141*(3), 439–443.

Kalashnikova, M., Peter, V., Di Liberto, G. M., Lalor, E. C., & Burnham, D. (2018). Infant-directed speech facilitates seven-month-old infants' cortical tracking of speech. *Sci Rep, 8*(1), 13745. doi:10.1038/s41598-018-32150-6

Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences, 100*(15), 9085-9089.

Kimura, M., Schröger, E., Czigler, I., & Ohira, H. (2010). Human visual system automatically encodes sequential regularities of discrete events. *Journal of Cognitive Neuroscience, 22*(6), 1124-1139.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences of the United States of America, 97*(22), 11850-11857. doi:DOI 10.1073/pnas.97.22.11850

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology, 62*, 621-647.

Kwok, V., Niu, Z., Kay, P., Zhou, K., Mo, L., Jin, Z., . . . Tan, L. H. (2011). Learning new color names produces rapid increase in gray matter in the intact adult human cortex. *Proceedings of the National Academy of Sciences of the United States of America, 108*(16), 6686-6688. doi:10.1073/pnas.1103217108

Lemhöfer, K., Spalek, K., & Schriefers, H. (2008). Cross-language effects of grammatical gender in bilingual word recognition and production. *Journal of Memory and Language, 59*(3), 312-330.

Liu, Q., Li, H., Campos, J. L., Teeter, C., Tao, W., Zhang, Q., & Sun, H. J. (2010). Language suppression effects on the categorical perception of colour as evidenced through ERPs. *Biol Psychol, 85*(1), 45-52. doi:10.1016/j.biopsycho.2010.05.001

Liu, Q., Li, H., Campos, J. L., Wang, Q., Zhang, Y., Qiu, J., . . . Sun, H. J. (2009). The N2pc component in ERP and the lateralization effect of language on color perception. *Neurosci Lett, 454*(1), 58-61. doi:10.1016/j.neulet.2009.02.045

Lucy, J. A. (1997). Linguistic relativity. *Annual review of anthropology, 26*(1), 291-312.

Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Frontiers in Psychology, 3*, 54.
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.
Lupyan, G., & Mirman, D. (2013). Linking language and categorization: Evidence from aphasia. *Cortex, 49*(5), 1187-1194.
Lupyan, G., Mirman, D., Hamilton, R., & Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition, 124*(1), 36-49. doi:10.1016/j.cognition.2012.04.002
MacSweeney, M., Woll, B., Campbell, R., Calvert, G. A., McGuire, P. K., David, A. S., . . . Brammer, M. J. (2002). Neural correlates of British sign language comprehension: spatial processing demands of topographic language. *Journal of Cognitive Neuroscience, 14*(7), 1064-1075.
Mahon, B. Z., & Hickok, G. (2018). Arguments about the nature of concepts: Symbols, embodiment, and beyond. *Psychonomic Bulletin & Review*, *23*, 941–958.
Maier, M., Glage, P., Hohlfeld, A., & Abdel Rahman, R. (2014). Does the semantic content of verbal categories influence categorical perception? An ERP study. *Brain and Cognition, 91*, 1-10. doi:10.1016/j.bandc.2014.07.008
Maier, M., & Abdel Rahman, R. (2018). Native language promotes access to visual consciousness. *Psychological Science,* 29, 1757-1772.
McClelland, J. L. (2006). How far can you go with Hebbian learning, and when does it lead you astray. *Processes of change in brain and cognitive development: Attention and performance xxi, 21*, 33-69.
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. doi:10.1073/pnas.1111860108
Newman, A. J., Supalla, T., Hauser, P. C., Newport, E. L., & Bavelier, D. (2010). Prosodic and narrative processing in American Sign Language: An fMRI study. *NeuroImage, 52*(2), 669-676.
Nitsche, M. A., & Paulus, W. (2000). Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. *The Journal of physiology, 527*(3), 633-639.
Ocelák, R. (2016). "Categorical perception" and linguistic categorization of color. *Review of Philosophy and Psychology*, *7*, 55–70.
Oxbury, J. M., Oxbury, S. M., & Humphrey, N. K. (1969). Varieties of colour anomia. *Brain, 92*(4), 847-860.
Paluy, Y., Gilbert, A. L., Baldo, J. V., Dronkers, N. F., & Ivry, R. B. (2011). Aphasic patients exhibit a reversal of hemispheric asymmetries in categorical color discrimination. *Brain and Language, 116*(3), 151-156. doi:10.1016/j.bandl.2010.11.005
Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108*(1), 155-184.
Pavlenko, A., & Malt, B. C. (2011). Kitchen Russian: Cross-linguistic differences and first-language object naming by Russian–English bilinguals. *Bilingualism: Language and Cognition, 14*(1), 19-45.
Perry, L. K., & Lupyan, G. (2014). The role of language in multi-dimensional categorization: evidence from transcranial direct current stimulation and exposure to verbal labels. *Brain Lang, 135*, 66-72. doi:10.1016/j.bandl.2014.05.005
Perry, L. K., & Lupyan, G. (2017). Clarifying the label–categorisation link. *Language, Cognition and Neuroscience, 32*(8), 950-953. doi:10.1080/23273798.2017.1328124

Pinker, S. (1995). *The language instinct: The new science of language and mind* (Vol. 7529): Penguin UK.
Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*: Penguin. New York.
Pulvermüller, F. (2018). Neural reuse of action perception circuits for language, concepts and communication. *Progress in neurobiology, 160*, 1-44.
Rabovsky, M., Sommer, W., & Abdel Rahman, R. (2012). Depth of conceptual knowledge modulates visual processes during word reading. *Journal of Cognitive Neuroscience, 24*(4), 990-1005.
Rentzeperis, I., Nikolaev, A. R., Kiper, D. C., & van Leeuwen, C. (2014). Distributed processing of color and form in the visual cortex. *Frontiers in Psychology, 5*, 932.
Roberson, D., Davidoff, J., & Braisby, N. (1999). Similarity and categorisation: Neuropsychological evidence for a dissociation in explicit categorisation tasks. *Cognition, 71*(1), 1-42.
Roberson, D., Davidoff, J., Davies, I. R., & Shapiro, L. R. (2004). The Development of Color Categories in Two languages: a longitudinal study. *Journal of Experimental Psychology: General, 133*(4), 554.
Roberson, D., Davidoff, J., Davies, I. R., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology, 50*(4), 378-411.
Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, *107*, 752-762.
Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928.
Siok, W. T., Kay, P., Wang, W. S., Chan, A. H., Chen, L., Luke, K.-K., & Tan, L. H. (2009). Language regions of brain are operative in color perception. *Proceedings of the National Academy of Sciences, 106*(20), 8140-8145.
Sonty, S. P., Mesulam, M. M., Thompson, C. K., Johnson, N. A., Weintraub, S., Parrish, T. B., & Gitelman, D. R. (2003). Primary progressive aphasia: PPA and the language network. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society, 53*(1), 35-49.
Tan, L. H., Chan, A. H., Kay, P., Khong, P.-L., Yip, L. K., & Luke, K.-K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Sciences*, *105*(10), 4004-4009.
Thierry, G. (2016). Neurolinguistic Relativity: How Language Flexes Human Perception and Cognition. *Lang Learn, 66*(3), 690-713. doi:10.1111/lang.12186
Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences, 106*(11), 4567-4570. doi:10.1073/pnas.0811155106
Trueswell, J. C., & Papafragou, A. (2010). Perceiving and remembering events cross-linguistically: Evidence from dual-task paradigms. *Journal of Memory and Language, 63*(1), 64-82.
von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance.
von Stutterheim, C., & Carroll, M. (2006). The impact of grammatical temporal categories on ultimate attainment in L2 learning. *Educating for advanced foreign language capacities: Constructs, curriculum, instruction, assessment*, 40-53.

von Stutterheim, C., & Nuse, R. (2003). Processes of conceptualization in language production: Language-specific perspectives and event construal. *Linguistics*, *41*(5; ISSU 387), 851-882.
Wenger, E., Brozzoli, C., Lindenberger, U., & Lövdén, M. (2017). Expansion and renormalization of human brain structure during skill acquisition. *Trends in Cognitive Sciences*, 21(12), 930-939.
Whorf, B. L. (1956). *Language, thought, and reality; selected writings*. Cambridge: Technology Press of Massachusetts Institute of Technology.
Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences, 104*(19), 7780-7785.
Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision*, *11*, 16.
Xia, T., Xu, G., & Mo, L. (2019). Bi-lateralized Whorfian effect in color perception: Evidence from Chinese Sign Language. *Journal of Neurolinguistics, 49*, 189-201. doi:10.1016/j.jneuroling.2018.07.004
Yang, J., Kanazawa, S., Yamaguchi, M. K., & Kuriki, I. (2016). Cortical response to categorical color perception in infants investigated by near-infrared spectroscopy. *Proc Natl Acad Sci U S A*, *113*(9), 2370-2375. doi:10.1073/pnas.1512044113
Yu, M., Mo, C., Zeng, T., Zhao, S., & Mo, L. (2017). Short-term trained lexical categories affect preattentive shape perception: Evidence from vMMN. *Psychophysiology, 54*(3), 462-468. doi:10.1111/psyp.12797
Zhong, W., Li, Y., Li, P., Xu, G., & Mo, L. (2015). Short-term trained lexical categories produce preattentive categorical perception of color: evidence from ERPs. *Psychophysiology*, *52*(1), 98-106. doi:10.1111/psyp.12294
Zhou, K., Mo, L., Kay, P., Kwok, V. P., Ip, T. N., & 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.