Perceiving colour through a language lens: a scoping review of experimental work on effects of language on colour perception

Owen Kapelle[1,2] (iD) and Monique Flecken[1]

[1]Faculty of Humanities, University of Amsterdam, Amsterdam, Netherlands and [2]Data Science Centre, University of Amsterdam, Amsterdam, Netherlands
Corresponding author: Owen Kapelle; Email: o.e.j.kapelle@uva.nl

Abstract

The popularity of colour perception as a vehicle to investigate language–perception interactions has led to a large body of experimental work. Recently, studies have focused on investigating the underlying cognitive and neural mechanisms of effects of language on colour perception. Because of substantial variation in experimental designs and the study conditions in these designs, evaluating and comparing the evidence reported in these studies remains complex. This is problematic, because language–perception interactions manifest themselves differently across cognitive contexts and task designs. To shed light on the precise conditions under which such effects are obtained, we conducted a scoping review on 72 experimental papers, and we assessed the experimental approaches taken. Based on this review, we recommend committing to an interdisciplinary approach, relying on knowledge of the neuroscience of perception. We provide specific examples of how future research can carefully investigate the relationship between cognitive load, attention, working memory and verbal label access.

Keywords: colour perception; language and perception; scoping review; verbal interference

1. Introduction

There has been a wide range of research exploring the question whether and in what ways language shapes our thought. The traditional debate on how language and thought interact is centred around two opposing views: universalism and relativism. The universalism view holds that variation between languages is limited, because languages are a product of human cognition, and the extent to which languages can vary is tied to the structure and boundaries of cognition (Regier et al., 2010). In this view, cognitive reality constrains linguistic variation. In contrast, the relativist view holds that linguistic systems can vary, as languages are based on a community's

conventions and needs. These differences between languages are theorised to affect a speaker's cognition. In turn, cognitive processes between speakers of different languages differ, since cognitive reality is reshaped to align with linguistic variation (Wolff & Holmes, 2011). The crucial difference between these two views is that the universalism view assumes cognition cannot be altered by language, while the relativism view assumes that languages do change the structure of cognition.

To study how and whether language can alter thinking, because of the difficulty in studying cognition as a whole, a large number of studies have focused on investigating the effect that language has on a specific cognitive process, that is, perception. These investigations have resulted in conclusions supportive of both the relativist and universalist views (see overview in Lupyan et al., 2020; Regier et al., 2010). In a seminal theoretical paper, Lupyan (2012) attempts to explain these opposing results and posits that the influence of language on perception is transient in nature; language effects on perception may be observed in one context, but remain absent in another. For example, some experimental tasks may encourage cognitive processing in a way that language effects are more evident than in other tasks. Because of this proposed dependency on context, future investigations should concentrate on investigating which experimental manipulations show an influence of language on perception, and which do not, to elucidate which task settings and contexts yield consistent language effects. For example, taking into account different cognitive processing demands in experimental tasks may shed light on the cognitive contexts in which language interacts with perception. The identification of tasks more prone to eliciting substantial linguistic effects facilitates a more precise exploration of the fundamental mechanisms underlying those effects, while also potentially offering insights into the context-dependent cognitive processes that are predisposed to elicit language-mediated effects on perception.

In line with the intention to better understand how language–perception interactions present itself in experimental context, recent reviews (e.g., Athanasopoulos & Casaponsa, 2020; Lupyan et al., 2020) expressed the need for a meta-analysis of the main findings to better understand what we can conclude from more than 70 years of study on language and colour perception. The work ranged from studies such as Brown and Lenneberg (1954), who looked at the relation between codability of colour and colour recognition, as well as Berlin and Kay (1969) who looked at basic colour terms, cross-linguistic work such as Roberson et al. (2005), who compared the colour-space of English speaking participants to speakers of Himba, up to Casaponsa et al. (2024) who looked at the stages of attention in which linguistic labels affect encoding in participants' native versus non-native language. While numerous reviews have examined the evidence concerning the interaction between language and colour perception (Athanasopoulos & Casaponsa, 2020; Lupyan, 2012; Lupyan et al., 2020; Regier & Kay, 2009; Roberson, 2005; Wright, 2014), none of these reviews employed a systematic approach in selecting the papers for review, and so a thorough systematic overview of the work done to date is lacking. Yet, it is essential to have a good overview of all the available evidence that is not selective or accidentally biased (Munn et al., 2018), because a meta-analysis requires a predetermined question that serves a single, clear purpose and without a systematic overview of previous work it is challenging to formulate such an objective question, especially because of substantial variation in experimental designs in which the underlying mechanistic differences even between two experimental conditions are not always clear (Akbiyik et al., 2022).

Our main focus in this scoping review will be on investigating which types of tasks are consistently employed to explore language effects on *colour* perception, and identifying the experimental context and cognitive processing demands which modulate language–perception effects. We will first give a theoretical background to the field, and then address three research questions:

1. Which research questions, regarding the role of language in colour perception, have been addressed?
2. What are the specific experimental designs and paradigms used?
3. Which task demands and cognitive mechanisms seem to mediate the role of language in colour perception?

Furthermore, we will present a case study on one of the more popular theoretically relevant approaches that have recently received some attention, the verbal interference paradigm. In this paradigm, the blocking of access to linguistic colour labels is achieved, and this is done to shed light on the necessity of linguistic retrieval for colour discrimination. Whilst in general, empirical studies employing verbal interference report reduced categorical colour perception – suggesting a prominent role for the linguistic labels during colour perception – some recent work shows contradictory findings, especially in bilingual participants (e.g., Nedergaard et al., 2023). A closer inspection of this paradigm and its role in shedding light on language–perception interaction allows us to elucidate certain methodological factors that can influence the study of language and colour perception. Ultimately, based on the analysis of the methodologies used and the discussion of the verbal interference paradigm, we make suggestions for future research on this topic to improve experimental approaches.

2. Background

2.1. The state of the art

Berlin and Kay (1969) proposed that colour terms in all languages centre around a universal structure of eleven 'basic colour categories' (Berlin & Kay, 1969), which have been influential in later work (Witzel, 2019). The theory of Berlin and Kay (1969) opposes earlier work by for example Gleason (Gleason Jr., 1955, in Berlin & Kay, 1969), who claimed that colour terms are ascribed 'arbitrarily' across languages. Berlin & Kay's work was followed up by more cross-linguistic investigations (e.g., Roberson et al., 2000) showing that a number of small languages did not follow this universal colour space proposed by Berlin and Kay: There were substantial differences between languages with regards to the cognitive structure of colour categories (Roberson et al., 2000; Roberson et al., 2005). This line of work sparked a tradition of research in which the effect of such linguistic differences on the perception of colours was examined, bringing the debate back to the relativistic views that were opposed by Berlin and Kay (1969).

In recent years, the debate on language–cognition interactions has aimed to recentre itself into a nuanced, middle-ground position (Regier et al., 2010; Regier & Kay, 2009; Wolff & Holmes, 2011) with a focus on *how* language influences perception (Lupyan, 2012). A more physiological approach, such as various neuro-imaging techniques, is recommended for this type of question because such an

approach can directly track the mechanisms underlying language effects on perception to elucidate the source of language effects on colour perception (Athanasopoulos & Casaponsa, 2020. Generally, recent studies on colour specifically investigated the effect of language on categorical perception.

Categorical perception of colour (CCP) is a derivative of the more general term categorical perception (CP). It was traditionally considered in the perception of speech (auditory perception), but it can also be extended to other modalities of perception (Harnad, 2003). CP is characterised by the decrease of a perceived distinction for exemplars or items that belong to the same category (within-category exemplars), and the perceived increase of between-category differences (Harnad, 2003). Categorical perception is considered an important aspect of human perception, as it allows for efficient and fast categorisation of stimuli in the environment, that is, in speech perception our phonological categories allow us to distinguish /p/ from /b/ sounds and thus recognise words like /path/ and /bath/ as distinct. Categorical perception of colour (CCP) is observed in experimental settings as the overall better discrimination of between-category colours (e.g., green and blue) than within-category colours (e.g., two shades of blue). Colour perception is an interesting field of study because physically, colour is a continuous spectrum, but all languages have developed words with which these colours are categorised, thus creating 'cognitive' boundaries within the colour spectrum (Lupyan et al., 2020; Witzel, 2019). Berlin and Kay's basic colour terms (Berlin & Kay, 1969) are one (and certainly the most well-known) example of a way in which the colour spectrum can be divided into categories (Witzel, 2019).

The process of colour perception is generally thought of as comprising two stages: the physiological stage which takes place in the eyes, and the cognitive stage, which consists of mental processes which are not purely physical (Witzel, 2019). In this physiological stage of colour perception, the colour differences that can be distinguished depend on which wavelengths the cones of the human eye can detect (for an elaborate explanation, I refer the reader to Witzel (2019), p. 509). The eye can detect many more differences between colours than we generally experience or perceive consciously. To deal with the overload of incoming information, higher-order cognitive processes filter the perceived colour differences and categorises them. This is the core of what happens in the second stage of colour perception. These higher-order cognitive processes affect which colour differences are experienced and which ones are not, by adjusting our sensitivity to differences between colours that are relevant to us. Other subtle colour differences detected by the eyes are discarded (Witzel, 2019) and the colours that are actually experienced consciously are categorised and filtered by the higher-order cognitive processes. It is theorised that language, in the form of the colour labels present in a given language, determines which differences in colours are relevant to us and which are not. Precisely which labels can aid speakers in mapping the relevant colour differences can vary, and Witzel (2019) points out that the existence of specifically eleven terms according to the categories defined by Berlin and Kay (1969) must not be considered a rigorous upper limit. For example, an extensively studied colour naming model by Mylonas and MacDonald (2012) identified a colour space with 30 colour categories. In a condition of free colour naming they clearly see evidence for the existence of a category for turquoise, but in a condition involving constrained naming, they only allowed the 11 basic colour categories, this turquoise category was partly identified as green and partly as blue (Mylonas & MacDonald, 2012). This shows that people's

categorisation of colour space can be influenced by pre-supposed categories, such as the ones proposed by Berlin and Kay (1969). This is often referred to as 'the warping of colour space'.

2.2. *Mechanisms underlying language–perception interaction*

A central question in research on language–perception interaction has long concerned the 'strength' and scale with which language affects colour perception; whether the changes are long-lasting and irreversible, or whether they are short-term and context-dependent. In the most extreme end, during infant and child development, language can *overwrite* our 'prelinguistic' colour categories determined by physiological constraints of the eye in a long-lasting fashion (Wolff & Holmes, 2011), similar to how infants lose the ability to perceive phonetic categories not relevant to their native language (Werker & Tees, 1984). Alternatively, language may influence our perception of colour categories in a short-term, in the moment fashion (Lupyan, 2012). People speaking different languages indeed show differences in colour perception along the lines of their linguistic categories, but this effect appears to be less absolute than is the case in phoneme perception: Cross-linguistic differences in colour perception can be abolished by relatively simple task manipulations such as intended in verbal interference manipulations, in which the categorical perception of colour is influenced by manipulating whether participants can access a colour word or not (Lupyan, 2012). Blocking access to a colour word most often results in a lack of reliance on colour categories during perceptual tasks. However, in some cases, a participant's inability to access a colour label may instead result in increased or more efficient categorical perception, possibly because the loss of detail associated with access to the label would be detrimental to making a categorical distinction (Nedergaard et al., 2023). For instance, in Suegami and Michimata (2010), if the experimental design encouraged label usage, when participants needed to recognise the difference between two colours within a colour label category (e.g., 'blue'), blocking access to that label with verbal interference showed a facilitation effect in the within-category condition, suggesting that blocking the label enabled participants to make more precise categorical discriminations. González-Perelli et al. (2017) found a similar effect in Spanish participants, who showed increased categorical perception under verbal interference of the Spanish greater category label 'Azul' when they had to discriminate between what is for Uruguayan participants 'Azul' and 'Celeste'. Finally, verbal interference may also result in faster performance on a perceptual task than when there is no verbal interference, because of the time it takes to access a colour word that is used to categorise colours, when the categorisation process is similarly successful without accessing a word (Nedergaard et al., 2023).

The verbal interference paradigm is typically used to address the distinction between relativism and universalism in, for example, the processes underlying colour perception (see Section 5 below). In a seminal paper by Winawer et al. (2007), Russian speakers, who have two distinct labels for blue (e.g., light blue, 'goluboy' and dark blue, 'siniy') showed CCP along these two categories, while English speakers, with no such distinct categories, showed CCP along the lines of one category for blue. Russian and English participants thus showed a different perception of colour because of the

existence of different colour labels between the two languages, also suggesting a different structure of colour representation in Russian speakers' minds as compared with English. However, when these same Russian speakers were performing a perceptual discrimination task whilst at the same time having their ability to use linguistic labels during this task artificially compromised (as their language system was occupied with verbal interference manipulation), this effect disappeared. The disappearance of the effect suggests instead that there is no fundamentally different structuring of the underlying perceptual colour categories.

The simultaneous existence of a language-specific Russian colour category for light and dark blue and the relatively easy removal of the language-specific effect on the perception of blue in the study by Winawer et al. (2007) implies the existence of a blue category that is both specific to Russian, as well as non-specific, and, at least sometimes identical to the blue category for English speakers. Either the additional colour label that is present in Russian means that there is an 'additional' blue category compared with the basic category of blue, or language affects the process of colour perception itself – in the moment –, in that the underlying basic colour categories are differently recruited based on (linguistic) context. Lupyan (2012) argues that language and cognitive structures are inter-twined, and the exact role of language is dependent on the combination of the cognitive, visual and linguistic processes that are involved in colour categorisation during perceptual processing.

In explaining such in-the-moment effects of language on colour categorisation, it is important to note that the process of colour categorisation requires a gener-alisation over some aspect that is different between two physically non-identical stimuli, for example, two different shades of green, to regard these two colours both as being green. For example, using the word 'green' selectively activates some perceptual features that are diagnostic of this category making the difference between shades of green smaller or less salient. This is the top-down augmentation of perceptual representations by language (Lupyan, 2012). Visual representations can be up- or down-regulated by language so as to make two stimuli more or less different from each other. This is possible because, after learning that some colours are green while others are blue, the perceptual representations by a green-coloured object also activate the verbal label 'green' because of association. This results in a temporary 'warping' of colour space, where everything connected to 'green' is activated and generalised as being green. Thus, viewing something that is green activates the label 'green', which activates the representations also belonging to green (Lupyan, 2012). The previously memorised information about what types of colours can be called 'green' is then mixed with the visual input to perceive the colour. The label feedbacks to the visual system the types of visual information that have been accumulated over time as belonging to the category 'green', and this information is subsequently used to assess the visual input that is perceived in an efficient way: if the label green activated similar features as the features that are being observed, the observation can be categorised as green. According to the label-feedback hypothesis (Lupyan, 2012), colour labels aid the processing of colours because they allow for what is called a *top-down* processing strategy, as opposed to a *bottom-up* processing strategy.

Top-down processing is when we rely on our prior experience to make predictions about what we expect to see, while bottom-up processing is when we mainly rely on the visual input itself (de Lange et al., 2018). When we make predictions about

incoming visual input, the brain attempts to match that input with these top-down expectations to make sense of the data, and predictions can especially change what is perceived when perceptual input is ambiguous (de Lange et al., 2018). According to de Lange et al. (2018), top-down processing most likely occurs when expectations are reliable, and stimuli are ambiguous (e.g., when information is available and necessary to make sense of the visual input). However, bottom-up processing is more likely when there are weak or unreliable expectations about the type of visual input that will be encountered. A colour label can be regarded as a prediction, with the information related to that label exerting a top-down influence on the colour that is actually perceived.

Lupyan (2012) poses that when studies observe cross-linguistic differences, we are actually witnessing this difference between bottom-up and top-down processes. To return to the example of Winawer et al. (2007): if we have distinct light blue and dark blue categories, and we can reliably expect that we have to distinguish between light and dark blue, we can set our expectations in such a way that the information we have about what belongs to 'goluboy' or 'siniy' can exert top-down influences on what we actually see. We then do not necessarily need to process all dimensions of the stimulus as carefully as we would need to have done without this top-down modulation. However, when such a strategy is not possible, for example, because English does not have distinct categories for light or dark blue, or when a task prevents participants from accessing the information through the use of the label, as is the case with verbal interference studies, the stimulus can only be processed bottom-up, relying strongly on the perceptual input itself. Thus, the existence of a label in a language enables top-down processing, which results in more efficient perception of a colour as long as top-down predictions are possible in a given context.

Language effects on perception are highly context-dependent. In our example about the categorisation of light blue and dark blue above, it is clear that predictability is essential to shaping expectations (Pilling et al., 2003). If we are unable to set accurate predictions for which distinction is important, we are less likely to initiate a top-down processing strategy (De Lange et al., 2018). This context dependency has not been systematically implemented into research task designs, which may result in some task contexts encouraging or discouraging certain types of processing strategies. In a recent paper, Akbiyik et al. (2022) stated that it remains unclear which exact mechanisms are implicated by the different conditions in often-used behavioural research tasks. In this review we map the experimental designs and methods of relevant studies to more closely scrutinise which aspects of the methodology are especially likely to affect the consistency with which we find language–perception interactions.

3. Method

Our systematic search for papers is based on the PRISMA guidelines for scoping reviews (Page et al., 2021). Our protocol states that we are interested in papers reporting experimental studies involving a research question along the lines of '(How) does language influence the perception of colour?'. We excluded studies that explore only the differential organisation of colour space between languages but do not provide experimental investigations of linguistic influences on the perceptual processing of colour. Additionally, we include experimental studies that were

published since 2000, a predetermined cut-off points to limit the scale of the review. We only include papers with non-clinical adult populations as participants and no case studies. We only include papers that were published in English and in journals qualified as 'qualitative' using SJR rankings from *scimagojr*.com The SJR (Scimago Journal Rank) is a metric used to assess the relative importance and influence of academic journals. For this review, we only include papers that have a SJR ranking of Q1 or Q2 in at least one of three lists: Psychology, Neuroscience and Multidisciplinary. There is no specific 'linguistics' category in SJR rankings.

We identified four topics to be included in our search, namely: *Language*, *Perception*, *Colour* and *Experimental research*. We created four different search strings for each of these topics. Note that the databases we used (e.g., Web of Science Core Collection, PsycINFO and Linguistics and Language Behaviour Abstracts (LBA)) required different syntax for the search (see online supplement file 4). The search terms (regardless of database-specific syntax) were the following:

> *Language* AND Perception* AND Colo*ur OR Color Perception* AND Experimental Research**

3.1. Screening process

Our searches yielded 3697 results on 02 November 2022 (Figure 1). After duplication removal and removal of papers published before 2000, the remaining 2286 articles were screened using Rayyan.ai (Ouzzani et al., 2016) an online software facilitating article screening using a visual interface. We removed those articles that were completely irrelevant and unrelated to language or psychology based on titles. The remaining 313 papers were screened on titles and abstracts based on the protocol inclusion criteria (File 5). Ratings could be either 'included', 'excluded' or 'maybe'. Papers that were marked as 'maybe' often required full-paper assessment, for example because it was unclear whether language was actually the focus of the paper, or because it was unclear whether participants were adults. The ratings were done by three reviewers: two reviewers screened all article abstracts and titles, and 1 reviewer screened one-third of the abstracts. Rating was done completely blind between reviewers to limit inter-reviewer influence. Afterwards, reviewer reliability ratings were computed. This showed that 32% or 99 papers were marked as 'exclude' by all screeners and 17% or 53 of all papers were marked as 'include' by all screeners. A large portion of the papers were marked as 'maybe' by at least one reviewer. Most often, papers marked as 'maybe' by one researcher were either unclear about their population in the abstract, or the focus of the paper was not clearly on the role of language in perception. Additionally, screeners two and three were conservative in their rating as they were interns. Papers marked with 'maybe' were included for a second assessment by screener 1 (97 papers, 31%). A total of 65 papers (20%) were flagged as screener conflicts, this meant that two or more researchers had marked the papers differently. The conflicted ratings were discussed between the reviewers until a consensus was reached about the paper. In total, 19 papers were included after a second assessment of the papers. Finally, 72 papers were included in the review. The online repository shows an overview of all conflicted and maybe papers and their further assessment in file 6.
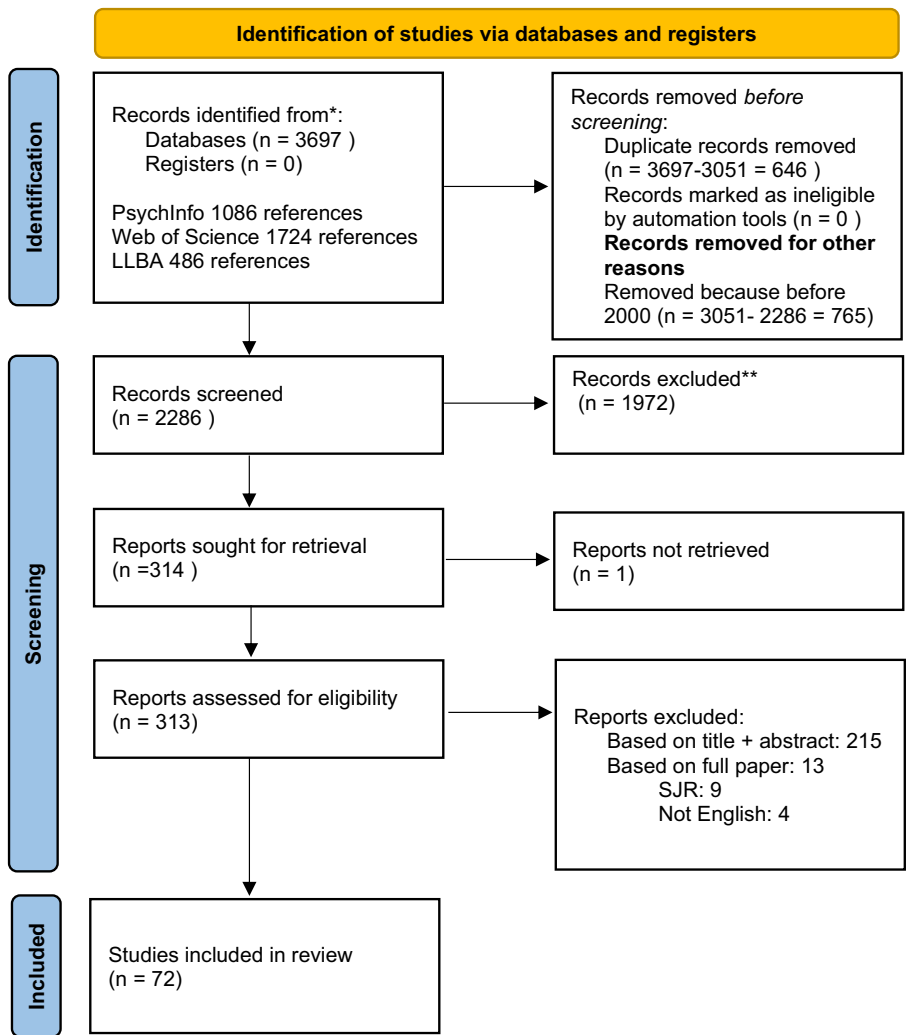
Figure 1. Review screening procedure (PRISMA).
Note: PRISMA diagram, as in Page et al. (2021). Diagram showing all stages of the screening process and the exact number of papers removed in each stage.

3.2. Data extraction

We extracted identificatory information from the papers such as the title, author, journal, year, SJR ratings and list and DOI, as well as the goal of the studies in that paper (such as research questions and conclusions), and more elaborate methodological information including the number of experiments in the paper, the number of participants tested per experiment, the colours tested, paradigms (i.e., verbal interference or visual field manipulations) and tasks used, whether a method used a behavioural or a physiological approach, and the dependent variables analysed.

The data repository in the online supplement additionally includes information on the categories in which we categorised the papers, as outlined in the next section.

4. Results

We used R Studio (R Studio Team, 2022) to process the data in this review, including the packages for tidy R coding (Wickham et al., 2019). For the figures included in this paper, we used ggplot (Wickham, 2016).

4.1. Categorisation of research questions in the literature

To get an overview of the research objectives of the included papers, we extracted the research questions and conclusions of all experiments reported (see data repository for details) and categorised them. We inferred three overarching research topics from the body of research, each with a number of sub-questions. Studies can be grouped into more than one topic. The first topic we identified is language specificity; these studies investigated to which extent language–perception effects can be attributed directly or specifically to language, and which to other cognitive factors or mechanisms. The second topic we identified is *neural mechanisms*; these studies investigated the underlying neural and cognitive mechanisms of a language–perception interaction. The third topic is *language experience*; studies on this topic investigated how a person's language experiences affect language–perception effects differently. Language experience can be operationalised as long-term experience, a contrast between native language and additional (second) language(s) within a speaker or a contrast between speakers of a different native languages, or as short-term experience, the outcome of a process of learning novel words (Table 1). In Figure 2, a distribution of the number of papers published with this topic over time is shown. The first research topic we further divided into four subtopics: The first subtopic we identified is *memory organisation*, in which studies address how colour words and the visual representation of a colour are connected to each other in visual or semantic memory. An example of this topic is study on the strength of the associations between colour words and the actual colour they represent (Twick & Cohen, 2011). The second topic concerns whether it is necessary to have *in-the-moment access to linguistic labels* for language–perception effects to arise, such as the verbal interference paradigm (e.g., in Winawer et al., 2007). The third sub-topic focuses on the *colour characteristics* that are sensitive to linguistic influences: For example, Agrillo and Roberson (2009) found that there was a larger effect of language on the perception of focal colours compared with non-focal colours, especially when they were easy to label. The fourth subtopic includes studies that investigated which *non-linguistic cognitive processes* can be

Table 1. The number of studies per topic

Research topic	Number of studies
1. Language specificity	49
2. Neural mechanisms	37
3. Language experience	20

Note: There were 72 unique studies in total. Studies can be ascribed to multiple research topics, thus, the total sum of studies per category is higher than 72. For example, if a study was categorised as being both about Language Specificity and Neural Mechanisms, it is counted once for each category, and thus counted twice in this table.
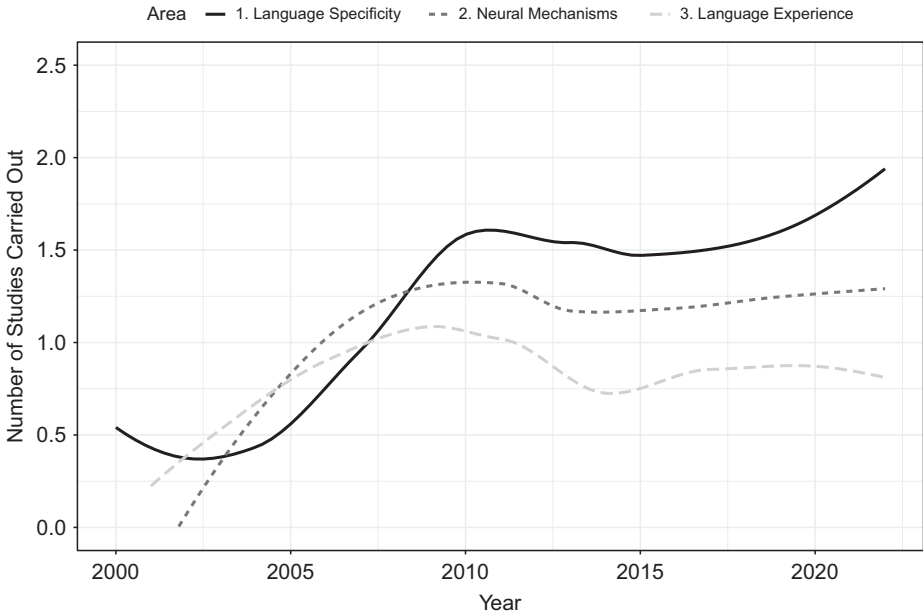
Figure 2. Studies conducted over time.
Note: The number of studies per category per year over time per category: Language Specificity (solid black), Neural Mechanisms (dashed dark grey) and Language Experience (dashed light grey).

identified as underlying the observed language–perception effects, for example, studies that investigate whether language effects may be largely attention-based (Štěpánková & Urbánek, 2021).

The second research topic focusses on the underlying neural mechanisms of language–perception effects with two sub-topics: The first sub-topic concerns the *stage of neural processing* during which linguistic influences arise. These studies investigate when, during the processing of colour, language plays a role, that is, during early or late processing stages (e.g., Forder et al., 2017). The second sub-topic concerns the *neural localisation of language effects on perception*, in which studies aim to localise the brain regions that appear to show different neural activations during colour perception when language is or is not involved. Studies that investigate this sometimes use a form of neuroimaging (e.g., Tan et al., 2008) or even neuro-stimulation (Akbiyik et al., 2022), although most studies take a behavioural approach (e.g., Gilbert et al., 2005.

The third research topic, also divided into two sub-topics, involves the specific extent of linguistic experience necessary to arrive at language–perception effects; either *long-term language experience* in terms of the native language that participants speak, or *short-term language experience*, in which participants learn new names for colour categories. In studies that investigate the role of short-term language experience, participants often undergo a training scheme in which they learn to divide a colour category into smaller categories by learning novel terms for these categories. For example, teaching participants new labels for (novel) colour categories to investigate whether a short-term language effect occurs for categories that were not known before training (Zhong et al., 2015). In the case of long-term language

experience, studies often make cross-linguistic contrasts between languages that differ in the labels that they have to describe colour. It is investigated whether these differences in linguistic categorisation influence colour perception (e.g., Roberson et al., 2000).

4.2. Methodological approaches

We mapped the various methodological approaches used and the distribution of these over the different research topics. The different aspects of methodologies we investigated are the type of techniques the studies employed (i.e., physiological or behavioural methods) and what type of task was used. Additionally, we did a more detailed analysis on one of the methodological approaches (the verbal interference paradigm, see Section 5) in which we also looked into the cognitive processes targeted by each task.

The majority of studies took a behavioural approach in their experiments (53 studies using a behavioural methods versus 19 studies using a physiological approach). When the types of approaches were compared between research question and topic (Figure 3), our analysis revealed that studies interested in the neural mechanisms underlying language–perception effects indeed utilise mostly physiological approaches, but that studies with different research topics almost exclusively use behavioural methods. While eighteen unique task designs were used in total, three tasks were most popular: the Alternative Forced choice task (21 studies), the visual search oddball task (23 studies, often with visual field manipulations) and the same-different task (13 studies) (Figure 4). Oddball tasks
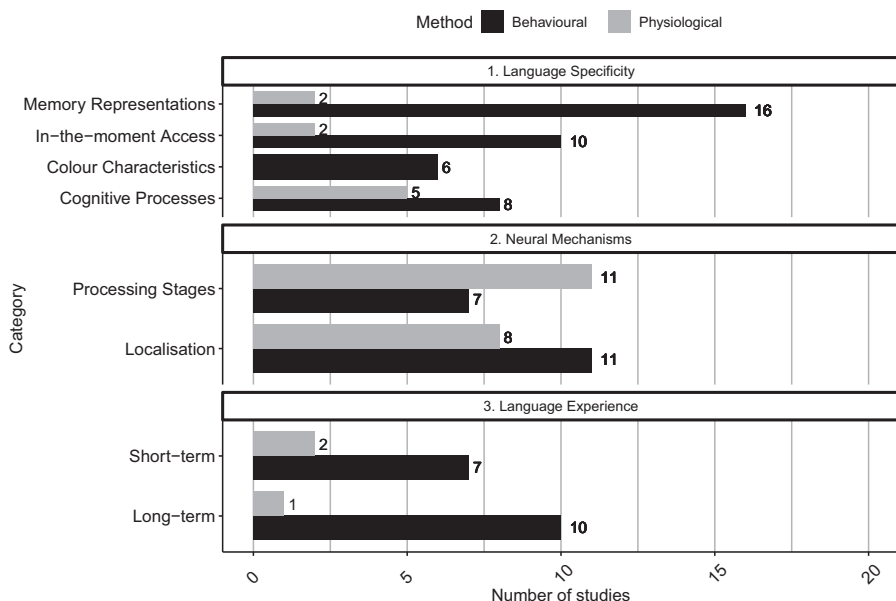

Figure 3. Number of studies per subtopic.
Note: Bar plot of the number of studies assigned to each sub-topic for each research topic, split on which type of method (Behavioural, dark grey or Physiological, light grey) was used in the study.
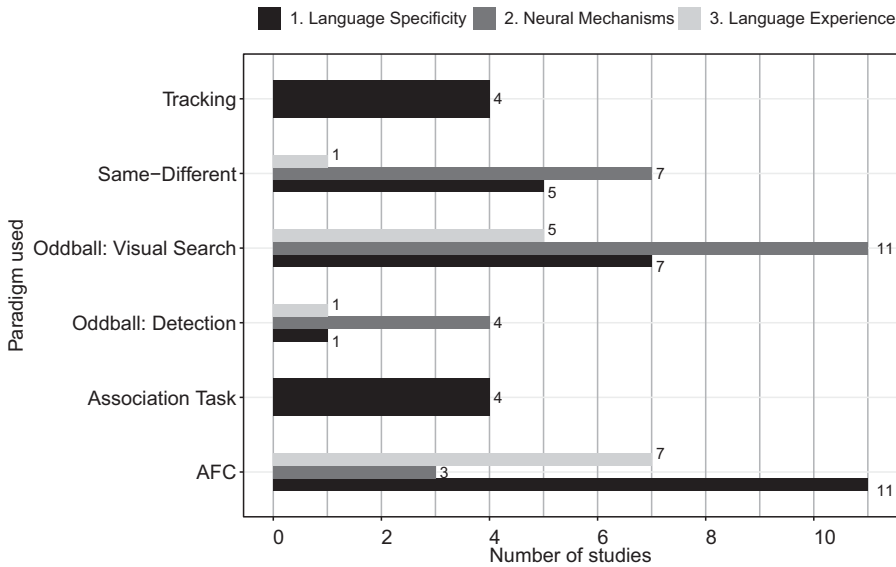
Figure 4. Tasks per category.
Note: Bar plot showing the number of studies using specific tasks per research topic in the sample, per category: Language Specificity (dark gray), Neural Mechanisms (middle gray) and Language Experience (light gray).

(both visual search and detection tasks) are used in the majority of studies interested in neural mechanisms, while the alternative forced choice, the most frequently used task in general, is used relatively little in papers interested in neural mechanisms (Figure 4).

5. Case study: verbal interference paradigm

To investigate more closely which task contexts mediate language–perception interactions, we conducted a case study on papers that employ a verbal interference paradigm. In this paradigm, participants are asked to overtly or silently reproduce and memorise a sequence of syllables, words, or digits, whilst performing various colour perception tasks. The function of this repetition component of the task is that it limits the possibility of immediate retrieval of a colour label, because the repetition task is saturating the language system. The assumption is that, because categorical perception of colours is because of people's reliance on automatic activation of colour, colours are easier to distinguish when they are encoded with distinct labels than when they do not have different labels (Lupyan, 2012). Access to the colour labels is necessary for such an advantage of the colour label. Impeding access to colour labels through the saturation of the language system with a verbal interference component is expected to diminish or, at least, reduce CCP. When a decrease in CCP is indeed consistently observed during verbal interference, it substantiates the notion that accessing the linguistic colour label in real-time is a fundamental prerequisite for CCP. However, such a decrease in CCP during verbal interference is, not always observed (e.g., Nedergaard et al., 2023). For example, when bilingual participants, whose first and second language differ in colour

category boundaries – such as blue categories in Lithuanian (two categories) and Norwegian (one category) – are tested in a colour discrimination task (Sinkeviciute et al., 2024), verbal interference appears to aid categorisation, but in a very specific manner: when verbal interference was provided in Lithuanian, but not in Norwegian, the bilinguals showed a colour category effect, that is, faster discrimination of between- as compared with within-category colours. This finding, and other findings showing language-specific modulation of interference effects (e.g., Athanasopoulos et al., 2015; Gonzalez-Perrilli, 2017; Nedergaard et al., 2023), suggest that real-time access to a colour label may not be a fundamental prerequisite for CCP in all contexts. Participant-related factors, such as their specific language experiences (in the case of bilingualism, which allows participants to utilise code-switching), as well as task-related contexts may modulate the influence of verbal interference.

5.1. *The theory behind verbal interference*

The verbal interference task aims to elicit what Baddeley calls 'articulatory suppression' (Baddeley, 2007; Henry, 2012). In Baddeley's working memory model, the component called the 'central executive' controls attention allocation and the dividing or switching of attention during a task. The central executive functions as a control system for working memory, while the storage of information happens in the 'phonological loop' or the 'visuospatial sketchpad' (in case of auditory or visual information respectively). The combination of multi-modal information and access to long-term memory happens in the 'episodic buffer'. Speech information is stored in the phonological loop, which consists of two subsystems: the phonological store, which can hold speech information up to two seconds, and the articulatory rehearsal system, which can be used to repeat information in the phonological store through verbal recoding. Verbal recoding of visual input (e.g., written language or pictorial stimuli) involves transforming that information into linguistic form (silently). Articulatory suppression (in case of verbal interference) happens when the articulatory rehearsal system in the phonological loop is blocked. When this happens, visual input cannot enter the phonological store (phonological coding is blocked) and verbal rehearsal of the contents of the phonological store is not possible. When the articulatory rehearsal mechanism is occupied with a task that requires attention, it becomes unavailable, and visual information cannot be recoded and stored verbally, and thus has to be stored visually (Henry, 2012). In the case of a verbal interference task during colour perception, the verbal interference stimulus modulates the accessibility of the colour label: less easy access to the colour label affects a participant's ability to distinguish two differently labelled colours because the distinction between these two colours cannot be remembered through the recoding of the colour to its label in the same manner. In this case study, we analysed how studies operationalised eliciting verbal interference because the observations of these studies rely on whether the specific operationalisation of blocking articulatory suppression is successful.

5.2. *General design*

A study's ability to successfully induce verbal interference (VI) is dependent on how much a given task engages the language system, as well as the demands of the (non-

linguistic) aspects of the study design that could potentially up- or downregulate possibly confounding cognitive processes. VI studies often use a dual-task approach, in which a primary perceptual task is paired with a secondary interference task that engages the language system. The use of a dual task increases attentional demands (Akbiyik et al., 2022). Furthermore, the central executive is responsible for allocating the appropriate amount of attention when and where required (Henry, 2012), thus, when a dual task is unbalanced in its requirements for each task, the central executive may allocate too much focus or attention to the dual task nature, losing capacity to focus on other aspects of a task. Differences in cognitive demands of the secondary task may affect the consistency with which possible effects, such as decreased CCP, are because of limited language access, or because of cognitive loads introduced by task design, making it difficult to distinguish between the role of language and other cognitive factors. To test whether the dual task nature of the interference task is responsible for the diminished CCP effects, VI studies often compare three conditions: linguistic interference, visual interference and no interference. The linguistic interference condition is expected to show a greater decrease in CCP than the other conditions, indicating that verbal interference affects the CCP more than the cognitive load or the attentional demands induced by the dual task. However, the control tasks for the verbal interference task vary across studies, which introduce some uncertainty about how these interference tasks affect the primary perceptual task. More details on the studies we analysed can be found are in the online supplement. All variables we analysed are visualised in Figure 5, which we will explain below.
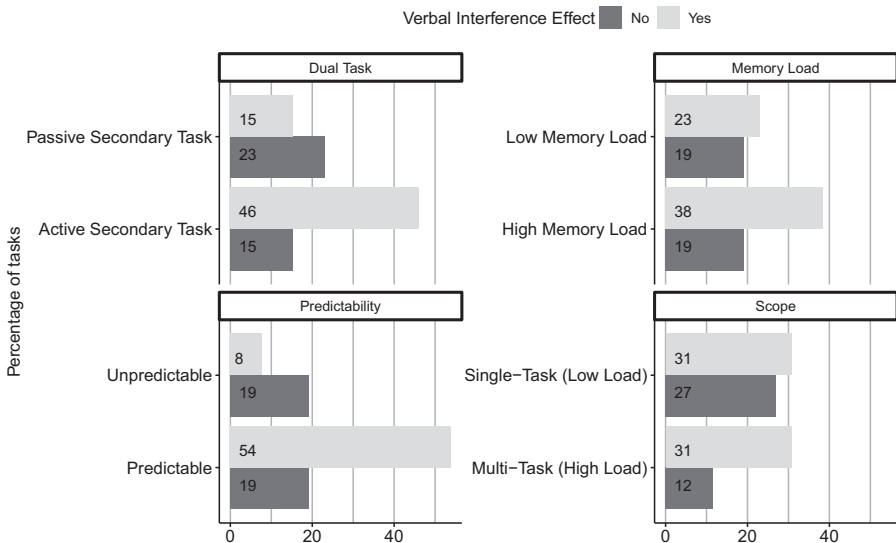

Figure 5. Task variables for verbal interference tasks.
Note: Bar plot with an overview of studies showing an interference effect (i.e., no CCP under verbal interference, compared with CCP under no interference conditions) per manipulation factor. Bar plot displaying the number of experiments that observe successful verbal interference effects (light grey) and studies that did not observe successful interference effects (dark grey).

5.3. Study designs in the sample

5.3.1. Primary task: manipulation of colour category

The primary task in most VI studies is the Alternative Forced Choice (AFC) task (Pilling et al., 2003; Roberson & Davidoff, 2000; Suegami & Michimata, 2010; Wiggett & Davies, 2008; Winawer et al., 2007). Other tasks include the same-different judgment task (Pilling et al., 2003; Akbiyik et al., 2022), and the visual search task (Gilbert et al., 2005; Akbiyik et al., 2022). These tasks differ in their reliance on label retention or visual differences detection, influenced by the presentation of target and distractor stimuli either simultaneously or consecutively, and the number of distractors presented. In the AFC paradigm, participants are presented with a target colour patch and two or more test patches, of which one is equal in colour to the target patch, and at least one is of another colour and functions as a distractor. Participants need to decide which of the test patches most closely matches the target. The visual search tasks require participants to identify a target among multiple distractors, presented on a screen simultaneously. The same-different task does not include a distractor, but requires a binary choice (same or not the same) between the target patch and only one single test-patch. In all tasks, the patches' colour differs either within the same category or between different categories.

Tasks with simultaneous stimulus presentation rely on categorisation and visual discrimination in the moment, with no information from memory consulted (Wright et al., 2015), which facilitates bottom-up processing. On the contrary, tasks with consecutive presentation depend more on maintaining colour information in memory (Pilling et al., 2003), and thus facilitate top-down processing. In the same-different task stimuli are presented consecutively, while in the visual search task stimuli from different categories are shown simultaneously (Wright et al., 2015). The consecutive presentation of stimuli in the AFC task functions to manipulate the memory resources required. This means that the same-different task relies most on memory retention and least on up-regulation of differences and down-regulating of similarities, the visual search task relies least on memory retention, and the AFC task is altered to be more memory-reliant or more focused on up- or down-regulating visual differences.

The choice between visual discrimination and memory reliance in a study can influence participants' inclination to use a 'labelling strategy' (Lupyan, 2012; Wright et al., 2015). A labelling strategy is hypothesised in tasks with high memory demands (Pilling et al., 2003), where participants use labels to remember categories instead of the exact colour observed. However, in tasks without memory demands, a labelling strategy may be irrelevant or even detrimental (Wright, 2014). Consequently, verbal interference may affect CCP differently in each task, depending on the necessity for visual discrimination and the extent of memory recruitment (Lupyan, 2012).

Comparing verbal interference studies, those with high memory load more consistently reported evidence for verbal interference on CCP, while those with low memory load showed inconsistent results. This inconsistency for studies with low colour memory load indicates that unknown variables are affecting the verbal interference effects instead of memory load or that the task is too easy, and there is a flooring effect. In instances with no memory requirements, CCP observations may be because of labels aiding visual discrimination rather than memory, which is potentially affected differently by verbal interference than categorical discrimination.

5.3.2. *Secondary task: interference manipulations*

5.3.2.1. Predictability. Pilling et al. (2003) theorise that when the type of interference stimulus within a series of trials is predictable (visual, verbal or no interfering stimulus), participants may change their task-strategy if their default strategy (such as a labelling strategy) is likely to not be successful. In such cases, any potential effects related to language will be diminished regardless of categories. Pilling et al. (2003) manipulated the predictability of the type of interference condition by employing a random-order design for an unpredictable condition, in which experimental blocks did not include only one type of interference but the interference types were mixed within each experimental block. The experiments with a predictable design instead ascribed one type of interference to an experimental block. They found that CCP indeed only survived verbal interference in unpredictable designs and conclude that a high level of predictability of whether the colour label will be available does not lead to articulatory suppression, but it instead discourages participants from taking a verbal strategy at all. Apart from Pilling and colleagues, only two other papers used an unpredictable study design: He et al. (2019), who combined it with a visual search task, and Suegami and Michimata (2010) who employed a similar 2AFC task. Suegami and Michimata (2010) were also unable to find diminished CCP in the case of verbal interference. In contrast, He et al. (2019), who employed a visual search task, did find diminished CCP under verbal interference with an unpredictable design. These different findings may be because of the memory requirements of the 2AFC task, employed by Pilling et al. (2003) and Suegami and Michimata (2010), compared with the visual search task, employed by He et al. (2019). Based on the theory by Pilling et al. (2003), participants may take a visual strategy to prevent label interference during a task altogether if they can predict when the interference condition will come. Since the likelihood that participants will take a visual strategy differs depending on the memory requirements of the task (Lupyan, 2012), the effect of predictability of verbal interference may interact with the amount of memory that is required for that task. If that is the case, a distinct pattern should emerge when comparing interference effects found in studies with high memory demands and low memory demands depending on the predictability of the verbal interference stimulus. With the current studies in this review, it is difficult to analyse this interaction between predictability and memory requirements, because these variables are not systematically varied. To further investigate the interaction, future experimental investigations should be employed with combinations of tasks that manipulate the predictability of the interference condition and compares the effect of this variable between a high memory and low memory task condition. Additionally, this interaction may prove more fruitful to analyse in a meta-review specifically tailored to the verbal interference task beyond just studies on colour perception.

5.3.2.2. Scope. Studies differed in whether the interference stimuli had single trial scope (an interfering stimulus that was only interfering for the length of one trial) or multi-trial scope (the interference stimulus required retention of the stimulus over multiple trials).Based on an analysis of the effects found in studies in this review, the majority of studies with multi-trial scope reported evidence of diminished CCP which was explained as being the result of verbal interference, while studies with single-trial scope reported such evidence in only half of the studies. It is likely that the additional working memory load introduced by a multi-trial interference stimulus

more consistently interferes with participant's ability to access the verbal label, but it may also be possible that this additional interference is not linguistic but a more general working memory effect. As digit memorisation is a task with generally high cognitive load (Sun & Zhang, 2022) it is possible that a task such as in Winawer et al. (2007) leans more heavily on the cognitive load aspect of the task rather than the verbal nature of it. It would be beneficial to systematically manipulate this within one experiment to explore the effect of a multi-trial and single-trial verbal interference stimulus on CCP and ideally, to include a measure of cognitive load.

5.3.2.3. Activity level. The fourth dimension that plays a role in the findings obtained in verbal interference paradigms is the activity level of the interference task, which affects how much attention is needed to perform the interference task (and thus how much can be allocated to the perceptual task). The different activity levels refer to whether or not the task requires an active response. An example of an active interference task is when participants have to read words out loud (Robinson & Davidoff, 2000) or recall a word that was shown to them before (Sun & Zhang, 2022). An example of a passive interference task is a stroop-interference task (Suegami & Michimata, 2010) in which participants first see a screen with either a colour word (verbal condition) or a number of crosses (non-verbal condition), and afterwards saw a screen on which they do a colour matching task.

Any contrast between a secondary task that requires an overt response, and one that does not, runs into the issue of differences in complexity and cognitive load (Pilling et al., 2003). The effect of differences in the amount of attention required by an active or passive task is not explicitly investigated by any of the verbal interference studies. The majority of studies with an active task, and thus high attention demands, found evidence for an interference effect. In contrast, for tasks with a passive secondary task, and thus with low attentional demands, a majority of the studies were not able to find evidence for a verbal interference effect on CCP. It is possible that the passive tasks did not succeed in an interference effect because not enough attention was paid to the interference stimulus, making the passive tasks less consistent in actually triggering verbal interference. It is also possible that the lack of diminished CCP was because of the lack of a dual task. This would mean that the dual task's cognitive load is necessary for verbal interference.

5.4. Discussion

In the case study of this review, we analysed studies utilizing a verbal interference paradigm, and identified a number of task design features that influenced the success of the verbal interference condition resulting in diminished CCP, some of which were not explicitly manipulated in the studies but were rather arbitrary task design features. Tasks demanding higher cognitive load were more successful in finding diminished CCP in verbal interference conditions. However, it remains difficult to elucidate precisely to what extent verbal interference is successful because it limits access to verbal labels, and to what extent the additional cognitive load makes the tasks inherently more difficult. This is because some of the factors that appeared to impact the success of a verbal interference condition were arbitrary results of a task design which were not the intended manipulations in most studies.

For example, there appears to be a possible interaction with the predictability of task contexts (because they encourage either a labelling strategy or a visual discrimination strategy) and the effect of memory requirements, but predictability was mostly not a systematic task manipulation, although its effect is consistent with the predictive coding theory (de Lange et al., 2018). We suggest future research to either systematically investigate this interaction experimentally or to conduct a meta-analysis on this interaction with a wide scope of research papers for inclusion. Another example is the scope of the interference stimulus, which affects memory load and working memory capacity, which is in line with predictions from the working memory model (Baddeley, 2007). Additionally, we identified factors that were indeed explicit manipulations in the studies reviewed, such as increased memory load because of item retention, or the activity level of a secondary task (affecting attention allocation). We conclude that task design and task context are crucial to whether a verbal interference stimulus succeeds in eliciting articulatory suppression. Furthermore, it is possible that the bilingual status of participants was also attributed to the unsuccessful verbal interference, because participants could code-switch between their two languages (e.g., Sinkeviciute et al., 2024). In the present research, it was not possible to take this variable into account because bilingual status was not reported on in most of the studies reviewed. Future research should focus on carefully balancing cognitive load in task design and consider how tasks promote verbal strategies, memory load, activity level, and attention allocation, and they may benefit from considering the theories mentioned above. Additionally, it is important to report all the languages that participants speak to control the effect of bilingualism (and other types of linguistic experiences) on verbal interference.

6. General discussion

We review the experimental tasks and research designs that have been used in previous research on the role of language in colour perception, to get a clear overview of the state of the art, and to understand which factors may play a role in obtaining language–perception interactions in this domain. We follow the label-feedback hypothesis by Lupyan (2012), which proposes that language effects on colour perception are context dependent because such influences arise in the moment, when processing a visual stimulus. Language thus functions as a means to set expectations for the category that a visual stimulus belongs to. Immediate context and task demands may have a significant effect on whether and how language is used during a perceptual task.

What stands out with regards to which types of research questions are addressed in studies is that, although cross-linguistic comparisons have traditionally been the main focus of research into the language–perception interaction, most of the literature in this review does not take this approach. Instead, while only 14 studies carry out cross-linguistic comparisons, more studies investigate the language specificity and the underlying neural mechanisms of language-on-perception effects (29 studies for both). This is partly line with the observation that a more neurobiological focus has emerged in recent years (Athanasopoulos & Casaponsa, 2020; Lupyan et al., 2020). However, the interest in the neural mechanism of the effect seems to have decreased in more recent years.

When considering the results concerning research questions and the paradigms used in studies collectively, we observe a discernible bias in paradigm selection based on the study's research topic. For example, while the alternative judgement task (AFC task) is a popular task, it sees limited use in studies investigating the neural mechanisms of CCP. Instead, it is used much more in the other two research topics. In contrast, the same-difference task, also a popular task, shows the opposite pattern: it is used most often in studies investigating neural mechanisms, and little in studies interested in different language experiences. The underlying reasons for this pattern remain unclear. Plausible explanations include that researchers interested in a certain topic simply adopt the type of experimental tasks and designs used in earlier studies, without any particular content-related motivation, or a lack of finding language effects in studies on a certain research topic when using a different paradigm, though this remains speculative.

Another observation regarding the alignment between task and research aim can be found in studies with the aim to localise the underlying neural mechanism of such effects in the brain. These studies almost exclusively make use of behavioural measures, mostly response times (e.g., in visual field manipulations). These types of measures do not directly observe the localisation of the neural mechanism underlying language–perception interaction in the brain. A likely explanation is that techniques that can be used to shed light on the neural machinery involved in language–perception effects, such as MEG or fMRI, bring substantial costs with them. This is also the case for studies which aim to investigate both the processing stages (temporal dimension) as well as the localisation of them: while there are a number of studies interested in both of these questions at the same time, none of them uses a technique that is sensitive to both these dimensions, as they carry the same disadvantage with regards to costs.

With regards to our third research question, we observe that the predictability of whether a trial will include verbal interference is the most important factor to take into account. The extent to which the demands of an experimental task are predictable can encourage or discourage the use of linguistic labels in participants. The importance of this is also in line with research on perception within a predictive coding framework (de Lange et al., 2018). The predictive coding framework assumes that the brain constructs an internal model of the world with which it encodes sensory inputs as parameters of this model. When perceptual input (e.g., a colour) is processed, this is interpreted against this model, so the difference between the predicted input (the expectation) and the actual input is sent forward to higher cognitive regions. With this process, perception is the minimisation of the error of a prediction, as judged against the input. This model decreases the amount of error that is allowed when perceptual input is complex, so there is higher reliance on the input itself and less on the higher-order prediction (de Lange et al., 2018). An example of a colour perception experiment in which this is relevant is a case when a colour to be discriminated is ambiguous, and the low predictability of the task context does not allow for participants to have robust expectations about what they are about to see. This low predictability discourages top-down processing of the percept. The task discourages the use of the colour label regardless of the condition that is actually encountered, to minimise the prediction error. Thus, when the same-different task requires a participant to match a second colour to an earlier colour, but it is unclear whether a verbal interference condition will limit verbal label access, participants may assume a visual strategy instead of a linguistic labelling strategy, just in case. This may

also result in an effect such as the one discussed in Nedergaard et al. (2023), in which participants were faster to discriminate colour in a condition *with* verbal interference: if the participants were not inclined to use a linguistic label at all, a verbal interference category may even enhance the ability to make a visual discrimination.

In future experimental work on this topic, it may prove beneficial to take into account the knowledge put forth by perception literature from cognitive science and neuroscience to allow us to form new hypotheses about the exact mechanisms underlying language–perception interactions (see also Slivac & Flecken, 2023). More specifically, we suggest taking more care to theoretically ground the choice of the task contexts within an experimental paradigm in an interdisciplinary manner. Cognitive and neuroscience studies already have much knowledge about the underlying mechanisms of (visual) perception. Therefore, we suggest studying more closely the work done on predictive coding and visual perception, and to consider language as a form of prior expectation (Slivac & Flecken, 2023).

For future research into the topic of how language affects our perception of colour, we suggest a large-scale study, comparing the different tasks, experimental designs, and manipulating the cognitive factors we review here, to shed light on the extent to which they mediate language–perception interaction. Such large-scale systematic collaborative studies have also been conducted in other fields which faced similar challenges. Collaborations such as this are a rising new way of doing research, often called 'big team science' (BTS) (Baumgartner et al., 2023). In BTS studies, researchers pool their knowledge and resources for a common goal, which increases the power, especially for complex questions, and increases the diversity of perspective involved in tackling such complex questions and allows for sharing of expertise and best practices (Baumgartner et al., 2023). Such a large-scale systematic comparison is specifically promising for research in psychology, as psychology research especially suffers from for example low power and thus lack of generalisability and failure in replicating results (Forscher et al., 2023). Although the current review does not focus on sample sizes, the average sample size of the studies in this review was relatively low (see file 1 on OSF), and most papers do not report statistical power. Therefore, this field of research into language and colour perception is especially fit for such an approach, although, for a more in-depth analysis of the power of studies in the field, a meta-analysis is required. The tasks that have been used in the past are relatively simple to conduct, and an international collaboration would be useful for the traditional cross-linguistic nature of the field. Another problem in the field, which this review cannot tackle, is related to the factors that have been studied but which did not show evidence of an influence on language–perception interaction. Since the current review only analyses published research on the topic of language–perception interactions, there remains the question of which studies and factors of interest are missing because of a publication bias. To map the mechanism underlying language–perception interactions, it is important to consider factors that are reliably reported as being influential, as well as those for which no reliable evidence could be gathered across experimental studies. Furthermore, we suggest studies to take into consideration the different factors we identified within one and the same study. As we indicated, many of these cognitive factors are not explicitly referred to in study designs, and therefore also not systematically manipulated. Only a handful of studies compared different task contexts across different tasks, such as Pilling et al. (2003) and Akbiyik et al. (2022), who both compared the 2AFC and same-different tasks' different memory reliance. We recommend studies to compare the cognitive load and

memory load required with the same task (e.g., an AFC task with the same number of distractors, only varying the amount of memory required). Such studies should also aim to use measures that can account for the involvement of distinct cognitive processes, that is, those associated with language processing and those linked to, for example, attentional processes not specifically involving the language system, as evidenced through the modulation of specific ERP components in experiment designs (e.g., see overview of the types of ERP components sensitive to manipulations in language–perception research in Athanasopoulos & Casaponsa, 2020).

In conclusion, our review suggests that the precise influence of a number of task design factors in experimental work on language-colour perception interactions should be more systematically investigated. The current literature exhibits considerable variability in task design and modulating factors, hindering definitive conclusions and complicating meta-analyses. Research has evolved beyond the basic inquiry into language's influence on colour perception, instead focusing on the underlying neural and cognitive mechanisms involved. This current trend would benefit from an inclusion of the appropriate neuro-physiological measurements. Additionally, linguistic influences on perception are intricately linked to the cognitive demands of the task. The implicit encouragement of language use, driven by participants' expectations for upcoming trials, is pivotal for linguistic effects on colour perception. For instance, tasks with predictable demands that encourages fast processing of stimuli may consistently induce language use and consequent effects. Such consistency is necessary to investigate the underlying cognitive and neural mechanisms with precision. Precise neuro-physiological measurements and task designs aligned with cognitive demands are crucial for advancing our understanding of how language shapes colour perception.

Data availability statement. We have made the datasets used for this review available. File 1 is a data repository of all coded information referred to in this review. File 2 is a data repository of the verbal interference paradigm reviewed files with coded information. File 3 is a more simple reference list of all papers reviewed. File 4 is our protocol inclusion criteria list. File 5 is our search protocol. File 6 is an overview of screeners assessment of papers. Files are accessible on Open Science Framework via the following link: DOI 10.17605/OSF.IO/BW2TS or https://osf.io/bw2ts/?view_only=cfc27053f8b2440ead0c1c86fb737c06.

Competing interest. The authors declare no competing interests exist.

References

Agrillo, C., & Roberson, D. (2009). Colour language and colour cognition: Brown and Lenneberg revisited. *Visual Cognition*, 17(3), 412–430. https://doi.org/10.1080/13506280802049247

Akbiyik, S., Göksun, T., & Balci, F. (2022). Elucidating the common basis for task-dependent differential manifestations of category advantage: A decision theoretic approach. *Cognitive Science*, 46(1). https://doi.org/10.1111/cogs.13078

Athanasopoulos, P., & Casaponsa, A. (2020). The Whorfian brain: Neuroscientific approaches to linguistic relativity. *Cognitive Neuropsychology*, 37(5–6), 393–412. https://doi.org/10.1080/02643294.2020.1769050

Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Riches, N., & Thierry, G. (2015). Two languages, two minds: Flexible cognitive processing driven by language of operation. *Psychological Science*, 26(4), 518–526. https://doi.org/10.1177/0956797614567509

Baddeley, A. (2007). *Working memory, Thought, and action*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528012.001.0001

Baumgartner, H. A., Alessandroni, N., Byers-Heinlein, K., Frank, M. C., Hamlin, J. K., Soderstrom, M., Voelkel, J. G., Willer, R., Yuen, F., & Coles, N. A. (2023). How to build up big team science: A practical

guide for large-scale collaborations. *Royal Society Open Science*, 10(6), 230235. https://doi.org/10.1098/rsos.230235

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. University of California Press.

Brown, R., & Lenneberg, E. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, 49, 454–462.

Casaponsa, A., García-Guerrero, M. A., Martinez, A., Ojeda, N., Thierry, G., & Athanasopoulos, P. (2024). Electrophysiological evidence for a whorfian double dissociation of categorical perception across two languages. *Language Learning*, 74(S1), 136–156. https://doi.org/10.1111/lang.12648

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

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. https://doi.org/10.1371/journal.pone.0178097

Forscher, P. S., Wagenmakers, E.-J., Coles, N. A., Silan, M. A., Dutra, N., Basnight-Brown, D., & IJzerman, H. (2023). The benefits, barriers, and risks of big-team science. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 18(3), 607–623. https://doi.org/10.1177/17456916221082970

Harnad, S. (2003). Categorical perception. In L. Nadel (Ed.), *Encyclopedia of cognitive science*. Nature Publishing Group, 67–4.

Gilbert, A., Regier, T., Kay, P., & Ivry, R. (2005). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), 489–494. https://doi.org/10.1073/pnas.0509868103

Gleason, Jr., H. A. (1955). *An introduction to descriptive linguistics, Rev. Ed.* New York: University of California Press.

González-Perelli, F., Rebello, I., Maiche, A., & Arévalo, A. (2017). Blues in two different Spanish-speaking populations. *Frontiers in Communication*, 2(18). https://doi.org/10.3389/fcomm.2017.00018

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

Henry, L. (2012). The development of working memory in children. The Working memory model, 1–34. https://doi.org/10.4135/9781446251348.n1

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

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

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, 18(1), 143. https://doi.org/10.1186/s12874-018-0611-x

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372, n160. https://doi.org/10.1136/bmj.n160

Pilling, M., Wiggett, A., Ozgen, E., & Davies, I. (2003). Is color 'categorical perception' really perceptual? *Memory & Cognition*, 31(4), 538–551. https://doi.org/10.1348/0007126042369820

Regier, T., Kay, P., Gilbert, A. L., & Ivry, R. B. (2010). Language and thought: Which side are you on, anyway? In B. Malt & P. Wolff (Eds.), *Words and the mind: How words capture human experience*, 165–182. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195311129.003.0009

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

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone-age culture. *Journal of Experimental Psychology. General*, 129(3), 369–398. https://doi.org/10.1037/0096-3445.129.3.369

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition*, 28(6), 977–986. https://doi.org/10.3758/BF03209345

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

Roberson, D. (2005). Color categories are culturally diverse in cognition as well as in language. *CROSS-CULTURAL RESEARCH*, 39(1), 56–71. https://doi.org/10.1177/1069397104267890

RStudio Team. (2022). *RStudio: Integrated development environment for R.* RStudio, PBC. http://www.rstudio.com/.

Sinkeviciute, A., Mayor, J., Vulchanova, M. D., & Kartushina, N. (2024). Active language modulates color perception in bilinguals. *Language Learning*, 74(S1), 40–71. https://doi.org/10.1111/lang.12645

Slivac, K., & Flecken, M. (2023). Linguistic priors for perception. *Topics in Cognitive Science*, 15(4), 657–661. https://doi.org/10.1111/tops.12672

Štěpánková, L., & Urbánek, T. (2021). Colour categorization and its effect on perception: A conceptual replication. *Journal of Psycholinguistic Research*, 52(1), 1–16. https://doi.org/10.1007/s10936-021-09791-2

Suegami, T., & Michimata, C. (2010). Effects of stroop interference on categorical perception in simultaneous color discrimination. *Perceptual and Motor Skills*, 110(3), 857–878. https://doi.org/10.2466/PMS.110.3.857-878

Sun, M., & Zhang, X. (2022). Language modulates categorical effects of moving color objects. *Perception*, 51(3), 210–217. https://doi.org/10.1177/03010066221078992

Tan, L., Chan, A., Kay, P., Khong, P., Yip, L., & Luke, K. (2008). Language affects patterns of brain activation associated with perceptual decision. *Proceedings of the National Academy of Sciences of the United States of America*, 105(10), 4004–4009. https://doi.org/10.1073/pnas.0800055105

Twick, M., & Cohen, A. (2011). Flexibility over automaticity: Separable representations for colours and words. *Visual Cognition*, 19(3), 392–414. https://doi.org/10.1080/13506285.2010.544463

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. https://doi.org/10.21105/joss.01686.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis.* Springer-Verlag. https://ggplot2.tidyverse.org.

Wiggett, A., & Davies, I. (2008). The effect of stroop interference on the categorical perception of color. *Memory & Cognition*, 36(2), 231–239. https://doi.org/10.3758/MC.36.2.231

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

Witzel, C. (2019). Misconceptions about colour categories. *Review of Philosophy and Psychology*, 10(3), 499–540. https://doi.org/10.1007/s13164-018-0404-5

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews. Cognitive Science*, 2(3), 253–265. https://doi.org/10.1002/wcs.104

Wright, O. (2014). Evidence from asymmetries in task performance: Colour category effects. In W. Anderson, C. P. Biggam, C. Hough, & C. Kay (Eds.), *Colour studies: A broad spectrum* (pp. 212–224). John Benjamins Publishing Company. https://doi.org/10.1075/z.191.14wri

Wright, O., Davies, I., & Franklin, A. (2015). Whorfian effects on colour memory are not reliable. *Quarterly Journal of Experimental Psychology*, 68(4), 745–758. https://doi.org/10.1080/17470218.2014.966123

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. https://doi.org/10.1111/psyp.12294